Formulations mathématiques et résolution numérique en mécanique

Christian Wielgosz, Bernard Peseux, Yves Lecointe

Formulations mathématiques et résolution numérique en mécanique

Christian Wielgosz

Bernard Peseux

Yves Lecointe

UNIVERSITÉ DE NANTES

Centrale Nantes

Polytech'Nantes

28/09/2004

Table des matières

INTRODUCTION GÉNÉRALE

La modélisation d'un grand nombre de situations d'intérêt pratique pour l'ingénieur ou chercheur, conduit à la recherche de solutions d'équations aux dérivées partielles, assorties de conditions aux limites et de conditions initiales, notamment en *Mécanique du Solide*, en *Mécanique des fluides*, en *Acoustique*, en *Thermique* ou en *Électromagnétisme*.

Ces équations sont posées en général sur des domaines géométriques, qui ne permettent pas l'emploi des techniques classiques de recherche de solutions exactes et elles doivent être résolues par des méthodes numériques et en particulier par la méthode des éléments finis, des équations intégrales (éléments de frontière, méthode des singularités) ou des différences finies.

L'étude des équations intégrales a commencé il y a plus d'un siècle sous la forme de la théorie du potentiel ou de l'identité de Somigliana par exemple, mais les développements concernant la résolution numérique ne datent que des années 1960.

Première partie

APPROXIMATION PAR RÉSIDUS PONDÉRÉS ET ÉLÉMENTS FINIS

Christian Wilegosz

Les développements des méthodes numériques reposent en fait sur des procédures d'automatisation du calcul des structures derrière lesquelles se « cachent » quelques opérations algébriques. Le but de la première partie du cours est de mettre en évidence la structure algébrique des problèmes d'équilibre de milieux continus. Après quelques rappels de dualité et des propriétés des espaces intervenant en mécanique, nous montrerons que les méthodes d'approximation peuvent se décomposer en deux groupes : approximation des équations aux déplacements (résidus pondérés, équations intégrales) et approximation des équations aux déformations ou aux contraintes (méthodes d'éléments finis de type déplacements ou forces).

Nous détaillons dans cette première partie la méthode des résidus pondérés et la méthode des éléments finis de type déplacements.

1

Théorie algébrique des milieux continus

1.1 Rappels sur la dualité en statique des milieux continus

Une partie importante de la mécanique des milieux déformables concerne l'étude du comportement de milieux soumis à des sollicitations appliquées infiniment lentement avec l'hypothèse des petites perturbations (HPP). Dans ce cas, et quelque soit le comportement du milieu, on peut faire apparaître quelques propriétés algébriques, très utiles grâce à leur généralité, des espaces qui interviennent en mécanique des milieux continus. Définissons tout d'abord ces espaces et les notations qui seront utilisées par la suite [Wielgosz 99]. On considère un milieu continu Ω de frontière $\partial\Omega$ et on note dS un élément de sa frontière. On note V l'espace vectoriel des champs de déplacements généralisés et Φ celui des charges généralisées. Ce couple d'espaces vectoriels est en dualité pour une forme bilinéaire $\langle\bullet,\bullet\rangle_1$ qui représente le travail des efforts extérieurs. Notons $\varphi=(f,t)$ le couple représentant l'ensemble des charges extérieures (charges de volume f et surfaciques t), et u le déplacement. Le travail des efforts extérieurs est défini par [1] :

$$(u,\phi)\in \mathrm{V}\times\Phi \quad\rightarrow\quad \langle u,\phi\rangle_1=\int_\Omega u\cdot f\,\mathrm{d}\Omega+\int_{\partial\Omega} u\cdot t\,\mathrm{dS} \tag{1.1}$$

On note E l'espace vectoriel des champs de déformations généralisées et Σ celui des champs de contraintes généralisées. Ce nouveau couple d'espaces vectoriels est en dualité pour une forme bilinéaire $\langle\bullet,\bullet\rangle_2$ qui représente le travail des contraintes dans les déformations, qui est l'opposé du travail des efforts intérieurs, et est défini comme suit :

$$(\varepsilon,\sigma)\in \mathrm{E}\times\Sigma \quad\rightarrow\quad \langle\varepsilon,\sigma\rangle_2=\int_\Omega \varepsilon:\sigma\,\mathrm{d}\Omega \tag{1.2}$$

On rappelle que la forme bilinéaire 1 représente le produit scalaire habituel $(\cdot)$ de deux vecteurs et que la forme bilinéaire 2 représente le produit doublement contracté $(:)$ entre deux tenseurs. L'opérateur déformation en HPP est l'opérateur gradient symétrique :

$$u\in \mathrm{V} \quad\rightarrow\quad \varepsilon=\mathrm{grad}_s\, u\in \mathrm{E} \tag{1.3}$$

et l'opérateur d'équilibre, transposé de l'opérateur déformation est tel que :

$$\sigma\in\Sigma \quad\rightarrow\quad (f,t)=(-\mathrm{div}\,\sigma;\sigma\cdot n)\in\Phi \tag{1.4}$$

1. On considère un problème de Neumann où les tensions sont imposées sur toute la frontière.

où div est l'opérateur divergence d'un tenseur. On peut résumer la situation à l'aide du schéma à quatre espaces suivant :

$$\begin{array}{ccc} \mathrm{E} & \langle\bullet,\bullet\rangle_2 & \Sigma \\ \mathrm{grad}_s \uparrow & & \downarrow(-\mathrm{div},\cdot n) \\ \mathrm{V} & \langle\bullet,\bullet\rangle_1 & \Phi \end{array} \tag{1.5}$$

et remarquer que l'expression du principe des travaux virtuels ne fait que transcrire la définition de l'opérateur transposé (ou encore adjoint) qui consiste à écrire l'égalité des deux formes bilinéaires :

$$\forall v \in \mathrm{V} \quad \langle v, \phi \rangle_1 = \langle \varepsilon, \sigma \rangle_2 \tag{1.6}$$

soit encore :

$$\forall v \in \mathrm{V} \quad \left\langle v, g^{\mathrm{T}}(\sigma) \right\rangle_1 = \langle g(v), \sigma \rangle_2 \tag{1.7}$$

où l'on a noté g, l'opérateur *déformation*, autrement dit le gradient et g^{T}, l'opérateur d'équilibre, c'est-à-dire la divergence au signe près.

1.2 Sous-espaces vectoriels remarquables

Il s'agit du sev de E des déformations compatibles et du sev de Σ des autocontraintes. Notons I le sev de E des déformations compatibles. Soit U le sev de V des déplacements cinématiquement admissibles :

$$\mathrm{I} = g(\mathrm{U}) \tag{1.8}$$

Définissons maintenant J, le sev de Σ des autocontraintes :

$$\mathrm{J} = \mathrm{Ker} g^{\mathrm{T}} \tag{1.9}$$

L'opérateur d'équilibre contient les conditions aux limites sur les forces par l'intermédiaire de la formule de Cauchy. Le sev des autocontraintes J, comme celui des déformations compatibles I, dépendra donc essentiellement des conditions aux limites (CL) sthéniques et cinématiques (par dualité).

Définissons maintenant l'orthogonalité entre ces deux sev remarquables I et J. Rappelons tout d'abord ce qu'on appelle polaire d'un sev d'un espace vectoriel en dualité avec un autre : si F et G sont deux espaces vectoriels en dualité pour une forme bilinéaire notée $\langle\bullet,\bullet\rangle$, et si M est un sev de F, on appelle polaire de M *dans* G, et on le note M^{o}, la partie de G définie par :

$$\mathrm{M}^{\mathrm{o}} = \{ y \in \mathrm{G} \setminus \forall x \in \mathrm{M},\ \langle x, y \rangle = 0 \} \tag{1.10}$$

C'est en fait l'orthogonal de M *dans* G. Pour bien le rappeler, résumons la situation par le petit schéma suivant :

$$\mathrm{M} \subset \mathrm{G} \quad \langle x, y \rangle \quad \mathrm{G} \supset \mathrm{M}^{\mathrm{o}} \tag{1.11}$$

En statique des milieux continus, cette propriété s'écrit :

$$\mathrm{J} = \mathrm{I}^{\mathrm{o}} \tag{1.12}$$

On dit que le sev des autocontraintes est le polaire (ou encore l'orthogonal) du sev des déformations compatibles et on résume la situation par :

$$\mathrm{I} \subset \mathrm{E} \quad \langle x, y \rangle_2 \quad \Sigma \supset \mathrm{J} = \mathrm{I}^{\mathrm{o}} \tag{1.13}$$

Si on admet maintenant le théorème des bipolaires, on a :

$$\mathrm{I}^{\mathrm{oo}} = \bar{\mathrm{I}} \tag{1.14}$$

où le symbole « barre » signifie « adhérence de ». On admettra, sans entrer dans des considérations topologiques qu'en général, en mécanique on a :

$$\mathrm{I}^{\mathrm{oo}} = \mathrm{J}^{\mathrm{o}} = \mathrm{I} \tag{1.15}$$

1.3 Structure algébrique en élasticité

On dit qu'un milieu solide est élastique linéaire, s'il existe un opérateur linéaire d de E dans Σ et on écrit :

$$\sigma = d : \varepsilon \tag{1.16}$$

L'opérateur d'élasticité d, tenseur du quatrième ordre reliant ε et σ, est un opérateur symétrique et défini positif. On a donc :

$$\begin{aligned} &\forall \varepsilon_1, \varepsilon_2 \quad \langle \varepsilon_1, d : \varepsilon_2 \rangle_2 = \langle \varepsilon_2, d : \varepsilon_1 \rangle_2 \\ &\forall \varepsilon \qquad \langle \varepsilon, d : \varepsilon \rangle_2 > 0 \end{aligned} \tag{1.17}$$

On complète le schéma à quatre espaces de la façon suivante :

$$\begin{array}{ccc} & \overset{d}{\rightarrow} & \\ \mathrm{E} & \langle \bullet, \bullet \rangle_2 & \Sigma \\ g \uparrow & & \downarrow g^{\mathrm{T}} \\ \mathrm{V} & \langle \bullet, \bullet \rangle_1 & \Phi \end{array} \tag{1.18}$$

Il est alors possible de construire l'opérateur qui résume les problèmes d'élasticité linéaire et de physique linéaire stationnaire ($\circ$ désigne la composition des applications) :

$$\mathscr{L}^* = g^{\mathrm{T}} \circ d \circ g \tag{1.19}$$

Illustrons ce schéma pour quelques problèmes de mécanique et de physique (et pour des problèmes de Neumann).

Pour une barre en traction où u représente le déplacement longitudinal, on a :

$$\begin{aligned} &-\mathrm{ES} u_{,xx} = p \quad +\mathrm{CL} \\ &g \equiv \bullet_{,x} \quad d \equiv \mathrm{ES} \quad g^{\mathrm{T}} \equiv (-\bullet_{,x} ; \text{e.n.}) \end{aligned} \tag{1.20}$$

et poutre en flexion, où u représente la flèche :

$$\begin{aligned} &\mathrm{EI}\,u_{,xxxx} = p \quad + \mathrm{CL} \\ &g \equiv \bullet_{,xx} \quad d \equiv \mathrm{EI} \quad g^{\mathrm{T}} \equiv (\bullet_{,xx}\,; \text{e.n.}) \end{aligned} \tag{1.21}$$

et l'opérateur déformation est l'opérateur dérivée première (ou dérivée seconde), c'est-à-dire le gradient en une dimension (ou le gradient du gradient) ; l'opérateur d'équilibre est au signe près l'opérateur dérivée première, c'est-à-dire la divergence en une dimension (ou la divergence de la divergence), et l'abréviation e.n. signifie équilibre des nœuds ; l'opérateur de comportement est réduit au scalaire ES (ou EI) et représente la rigidité à la traction (ou à la flexion). Nous sommes donc dans le cas où les espaces V, Φ, E et Σ sont des espaces de fonctions scalaires.

En deux dimensions (problèmes de potentiel ou de thermique où Δ est le laplacien scalaire et où u est le potentiel ou la température), on a :

$$\begin{aligned} &-\Delta u = f \quad + \mathrm{CL} \\ &g \equiv \mathrm{grad}\,\bullet \quad d \equiv i \quad g^{\mathrm{T}} \equiv (-\mathrm{div}\,\bullet ; \bullet \cdot n) \end{aligned} \tag{1.22}$$

L'opérateur déformation est le gradient, l'opérateur d'équilibre est composé de la divergence (au signe près) ainsi que du produit scalaire avec la normale extérieure n, et l'opérateur de comportement est l'identité. Nous sommes cette fois dans le cas où les espaces V et Φ sont des espaces de fonctions scalaires et les espaces E et Σ sont des espaces de fonctions vectorielles.

Pour un milieu continu en deux ou trois dimensions, où u est le vecteur déplacement, les opérateurs sont :

$$\begin{aligned} &-\Delta^{*} u = f \quad + \mathrm{CL} \\ &g \equiv \mathrm{grad}_s\,\bullet \quad d \equiv d_{ijkl} \quad g^{\mathrm{T}} \equiv (-\mathrm{div}\,\bullet ; \bullet \cdot n) \end{aligned} \tag{1.23}$$

où Δ^{*} représente l'opérateur de Navier. L'opérateur déformation est l'opérateur gradient symétrique (c'est en fait la partie linéaire du tenseur de Green-Lagrange), l'opérateur d'équilibre est composé de la divergence (au signe près) d'un tenseur et du produit scalaire de ce tenseur avec la normale extérieure, et l'opérateur de comportement d_{ijkl} est le tenseur élasticité du quatrième ordre. Nous sommes cette fois dans le cas où les espaces V et Φ sont des espaces de fonctions vectorielles et les espaces E et Σ sont des espaces de fonctions tensorielles.

Enfin pour des plaques en flexion, où u est la flèche, nous avons :

$$\begin{aligned} &\mathrm{D}(u_{,xxxx} + 2u_{,xxyy} + u_{,yyyy}) = p \quad + \mathrm{CL} \\ &g \equiv \overline{\overline{\mathrm{grad}}}\,\vec{\mathrm{grad}}\,\bullet \quad d \equiv d_p \quad g^{\mathrm{T}} \equiv (-\vec{\mathrm{div}}\,\overline{\overline{\mathrm{div}}}\,\bullet ; \bullet \cdot n) \end{aligned} \tag{1.24}$$

où D est le module de rigidité à la flexion des plaques, l'opérateur d'élasticité d_p est celui des plaques (contraintes planes intégrées dans l'épaisseur).

Les propriétés de linéarité de d, et l'orthogonalité de I et de J permettent ensuite de montrer que les structures de E et de Σ sont telles que :

$$\mathrm{E} = \mathrm{I} \oplus d^{-1}(\mathrm{J}) \qquad \Sigma = \mathrm{J} \oplus d(\mathrm{I}) \tag{1.25}$$

et on dit que J et d(I) (comme I et d^{-1}(J)) sont mutuellement polaires.

1.4 Résolution théorique du problème en déplacements

Le problème s'écrit directement avec l'opérateur raideur $\mathscr{L}^*$. Les efforts extérieurs $\varphi = (f,t)$ étant donnés, trouver le champ des déplacements u solution de :

$$\mathscr{L}^*(u) = \varphi \tag{1.26}$$

Précisons ce problème en distinguant l'opérateur d'équilibre local qui sera noté $\mathscr{L}$ et l'opérateur des conditions aux limites qui sera noté $\mathscr{B}$. Nous avons pour l'instant défini des conditions aux limites de type Neumann (équilibre de nœuds ou produit scalaire avec la normale extérieure) qui apparaissent dans l'opérateur d'équilibre. Il existe deux autres types de conditions aux limites : conditions de Dirichlet où le déplacement u est imposé sur toute la frontière $\partial\Omega$ du milieu, et conditions mixtes ou de Cauchy où le déplacement u est imposé sur une partie $\partial_1\Omega$ de la frontière et les tensions sont imposées sur l'autre partie $\partial_2\Omega$ de la frontière. Par exemple, un problème de Neumann s'écrit :

$$\begin{array}{l} \mathscr{L}(u) = f \quad \text{dans} \quad \Omega \\ \mathscr{B}(u) = t \quad \text{sur} \quad \partial\Omega \end{array} \tag{1.27}$$

Donnons maintenant quelques exemples de problèmes mixtes :

1. équilibre d'un barreau en traction encastré - libre de longueur l

$$\begin{array}{l} -\mathrm{ES}u_{,xx} = f \quad \text{sur} \quad x \in]0,l[\\ u(0) = 0 \\ \mathrm{ES}u_{,x}(l) = 0 \end{array} \tag{1.28}$$

2. équilibre d'une poutre console en flexion de longueur l

$$\begin{array}{l} \mathrm{EI}u_{,xxxx} = f \quad \text{sur} \quad x \in]0,l[\\ u(0) = u_{,x}(0) = 0 \\ \mathrm{EI}u_{,xx}(l) = -\mathrm{EI}u_{,xxx}(l) = 0 \end{array} \tag{1.29}$$

3. problème de potentiel

$$\begin{array}{lll} -\Delta u = f & \text{dans} & \Omega \\ u = u_o & \text{sur} & \partial_1\Omega \\ u_{,n} = t_o & \text{sur} & \partial_2\Omega \end{array} \tag{1.30}$$

4. problème d'élastostatique tridimensionnelle

$$\begin{array}{lll} \Delta^* u = f & \text{dans} & \Omega \\ u = u_o & \text{sur} & \partial_1\Omega \\ \sigma \cdot n = t_o & \text{sur} & \partial_2\Omega \end{array} \tag{1.31}$$

La solution théorique du problème en déplacements s'écrit :

$$u = \mathscr{L}^{*-1}(\varphi) \tag{1.32}$$

où $\mathscr{L}^{*-1}$ est une inverse à gauche de l'opérateur raideur.

En pratique, pour des problèmes en une dimension d'espace (barres, poutres), on sait construire cette inverse car il suffit de résoudre des équations différentielles (voir exemples 1 et 2). Pour des problèmes en deux ou trois dimensions, on ne sait pas trouver cette inverse car il faut alors résoudre un système d'équations aux dérivées partielles. Les approximations de ce problème se font par la méthode des résidus pondérés ou par la méthode des équations intégrales.

1.5 Résolution théorique en déformations et en contraintes

Le problème de mécanique des milieux déformables s'écrit cette fois : Les efforts extérieurs $\varphi = (f,t)$ étant donnés, trouver le champ des déplacements u, le tenseur des déformations ε et le tenseur des contraintes σ solutions des trois groupes d'équations suivantes :

– les équations de déformation que l'on résume par :

$$\varepsilon \in \mathrm{I} \tag{1.33}$$

– le principe des travaux virtuels :

$$\forall v \in \mathrm{V} \qquad \langle v, \phi \rangle_1 = \langle \varepsilon, \sigma \rangle_2 \tag{1.34}$$

– la loi de comportement d'élasticité linéaire :

$$\sigma = d : \varepsilon \tag{1.35}$$

Écrivons ce problème sous une forme plus condensée. Pour ce faire, supposons qu'il existe toujours un champ de contraintes particulier σ^* qui vérifie les équations d'équilibre. On peut donc écrire le PTV avec ce champ :

$$\forall v \in \mathrm{V} \qquad \langle v, \phi \rangle_1 = \langle \varepsilon, \sigma^* \rangle_2 \tag{1.36}$$

En faisant la différence avec l'expression précédente du principe, on obtient :

$$\forall v \in \mathrm{V} \qquad \langle \varepsilon, \sigma - \sigma^* \rangle_2 = 0 \tag{1.37}$$

Lorsque cette égalité est écrite avec des champs de déplacements cinématiquement admissibles, ε appartient à I et l'orthogonalité entre I et J prouve que $\sigma - \sigma^*$ est une autocontrainte. Donc, l'équilibre s'écrit :

$$\sigma \in \sigma^* + \mathrm{J} \tag{1.38}$$

et $\sigma^* + \mathrm{J}$ est appelé le sous-espace affine des contraintes statiquement admissibles. Regroupons maintenant les trois groupes d'équations sous la forme condensée :

$$\begin{aligned} &\varepsilon \in \mathrm{I} \\ &\sigma \in \sigma^* + \mathrm{J} \\ &\sigma = d\varepsilon \end{aligned} \tag{1.39}$$

La solution du problème en contraintes est :

$$\sigma \in d(\mathrm{I}) \cap (\sigma^* + \mathrm{J}) \tag{1.40}$$

et celle en déformations est :

$$\varepsilon \in \mathrm{I} \cap d^{-1}(\sigma^* + \mathrm{J}) \tag{1.41}$$

On peut montrer que la solution de chacun de ces problèmes existe et est unique. On trouvera une interprétation géométrique simple dans le plan des contraintes dans [Wielgosz 99]. La solution en contraintes est la projection de σ^* sur $d(\mathrm{I})$ parallèlement à J :

$$\sigma = \mathrm{proj}_{d(\mathrm{I})}\sigma^* \tag{1.42}$$

On a bien entendu une situation analogue dans l'espace des déformations : la solution en déformations est la projection de $d^{-1}(\sigma^*)$ sur I parallèlement à $d^{-1}(\mathrm{J})$:

$$\varepsilon = \mathrm{proj}_{\mathrm{I}} d^{-1}(\sigma^*) \tag{1.43}$$

Les approximations de la solution en contraintes conduisent à la méthode des éléments finis de type forces, celles de la solution en déformations à la méthode des éléments finis de type déplacements.

2

Méthode des résidus pondérés

2.1 Formulation

Notons u la solution exacte (même si on ne la connaît pas) de l'équation d'équilibre en déplacements :

$$\mathscr{L}^*(u) = \varphi \tag{2.1}$$

On définit le résidu comme étant l'erreur que l'on commet en remplaçant la solution exacte u par une approximation w du champ exact des déplacements u :

$$\mathscr{R}^*(w) = \mathscr{L}^*(w) - \varphi \tag{2.2}$$

Bien entendu ce résidu est nul pour u :

$$0 = \mathscr{R}^*(u) = \mathscr{L}^*(u) - \varphi \tag{2.3}$$

La forme variationnelle de cette relation s'écrit :

$$\forall v \in \mathrm{V} \qquad \langle v, 0 \rangle_1 = \langle v, \mathscr{R}^*(u) \rangle_1 = 0 \tag{2.4}$$

et v est appelée fonction de pondération ou encore fonction test. Si on est capable d'écrire cette formulation pour toute fonction test, alors les deux équations précédentes sont équivalentes. Soit maintenant w une approximation de u satisfaisant toutes les conditions aux limites, donc telle que (pour un problème de Neumann) :

$$\mathscr{B}(w) = t \tag{2.5}$$

La méthode des résidus pondérés consiste à annuler le résidu $\mathscr{R}(w) = \mathscr{L}(w) - f$ en écrivant la restriction de la formulation variationnelle précédente à l'opérateur $\mathscr{L}$:

$$\forall v \in \mathrm{V} \quad \langle v, \mathscr{R}(w) \rangle_1 = 0 \tag{2.6}$$

Supposons que l'on soit capable de trouver n fonctions w_i (vérifiant toutes les conditions aux limites) et écrivons w sous la forme :

$$w = \mathrm{W} \cdot \mathrm{X} \tag{2.7}$$

où W est un vecteur ligne contenant n fonctions w_i et X un vecteur colonne contenant les inconnues a_i du problème et choisissons n fonctions test v_i de manière à ce que l'on puisse écrire v sous une forme analogue :

$$v = \mathrm{V} \cdot \mathrm{P} \tag{2.8}$$

où V est un vecteur ligne contenant les fonctions v_i et P un vecteur colonne contenant des paramètres p_i. On a alors :

$$v^{\mathrm{T}} = \mathrm{P}^{\mathrm{T}} \cdot \mathrm{V}^{\mathrm{T}} \tag{2.9}$$

La forme intégrale précédente, avec la définition de la forme $\langle \bullet, \bullet \rangle_1$, s'écrit :

$$\forall \mathrm{P}^{\mathrm{T}} \qquad \int_\Omega \mathrm{P}^{\mathrm{T}} \mathrm{V}^{\mathrm{T}} \big(\mathscr{L}(\mathrm{W} \cdot \mathrm{X}) - f \big) \mathrm{d}\Omega = 0 \tag{2.10}$$

et conduit à un système matriciel de dimension $n \times n$:

$$\mathrm{K} \cdot \mathrm{X} = \mathrm{F} \tag{2.11}$$

où la matrice K est non symétrique, et F est un vecteur forces généralisées, tels que :

$$\mathrm{K} = \int_\Omega \mathrm{V}^{\mathrm{T}} \mathscr{L}(\mathrm{W}) \mathrm{d}\Omega \qquad \mathrm{F} = \int_\Omega \mathrm{V}^{\mathrm{T}} f \, \mathrm{d}\Omega \tag{2.12}$$

> On peut envisager d'utiliser plus de fonctions test m que de fonctions w_i $(m > n)$; dans ce cas le système matriciel est surdéterminé, et on peut le résoudre par approximation avec une méthode de moindres carrés.

Si l'on veut symétriser la matrice, on peut choisir les fonctions test v_i sur la même base que les fonctions w_i en écrivant :

$$v = \mathrm{W} \cdot \mathrm{P} \tag{2.13}$$

Dans ce cas la matrice K et le vecteur F deviennent :

$$\mathrm{K} = \int_\Omega \mathrm{W}^{\mathrm{T}} \mathscr{L}(\mathrm{W}) \mathrm{d}\Omega \qquad \mathrm{F} = \int_\Omega \mathrm{W}^{\mathrm{T}} f \, \mathrm{d}\Omega \tag{2.14}$$

et la matrice K est symétrique car l'opérateur $\mathscr{L}$ est symétrique (car $\mathscr{L}^* = g^{\mathrm{T}} \circ d \circ g$, d est symétrique, et toutes les conditions aux limites sont vérifiées par les w_i).

2.2 Exemples

2.2.1 Barreau bi-encastré soumis à une charge linéique

Avec une charge linéique f quelconque, le problème prend la forme :

$$\begin{aligned} &-\mathrm{ES} u_{,xx} = f \\ &u(0) = 0 \\ &u(l) = 0 \end{aligned} \tag{2.15}$$

et on peut en trouver la solution exacte u quel que soit le chargement f. Résolvons ce problème de manière approchée par résidus pondérés.

Approximation à un paramètre

Choisissons *une* approximation w du problème réel qui ne satisfasse que les conditions aux limites, c'est-à-dire :

$$w(0) = w(l) = 0 \tag{2.16}$$

On peut choisir n'importe quelle fonction satisfaisant ces conditions aux limites ; par exemple :

$$w(x) = ax(l-x) \tag{2.17}$$

Où a est une constante arbitraire que nous allons calculer par la méthode. Le résidu s'écrit :

$$\mathscr{R}(w) = -\mathrm{ES}w_{,xx} - f = 2\mathrm{ES}a - f \tag{2.18}$$

et la formulation résidus pondérés s'écrit, avec la définition de la forme bilinéaire 1 :

$$\forall v \in \mathrm{V} \qquad \int_0^l v(x)(2\mathrm{ES}a - f)\mathrm{d}x = 0 \tag{2.19}$$

Il reste à choisir la fonction test v. Comme tout est possible ($\forall$), choisissons la fonction la plus simple : v égale à la constante unité. On en déduit :

$$a = \frac{f}{2\mathrm{ES}} \tag{2.20}$$

La solution approchée du problème est donc :

$$v(x) = \frac{fx(l-x)}{2\mathrm{ES}} \tag{2.21}$$

On remarquera que c'est la solution exacte du problème du barreau lorsque la charge linéique f est constante, c'est-à-dire que la valeur de la constante a est proportionnelle à la valeur moyenne du chargement. Reprenons les calculs en conservant la même approximation w et en changeant de fonction test v. Pour éviter l'intégration, choisissons pour v une fonction de Dirac en milieu de barre, notée $\delta(l/2)$. La valeur de la constante est cette fois telle que :

$$\int_0^l \delta(l/2)(2\mathrm{ES}a - f)\mathrm{d}x = 0 \tag{2.22}$$

Elle est donc reliée à la valeur du chargement en milieu de barre. On appelle cette méthode la collocation par points, le point de collocation étant celui où on a placé la fonction de Dirac.

Approximation à deux paramètres

Choisissons une approximation à deux fonctions. Nous avons choisi une fonction parabolique précédemment ; nous aurions pu choisir une fonction circulaire ou autre chose, l'essentiel étant d'utiliser une fonction au moins deux fois dérivable pour pouvoir calculer le résidu. Résolvons le même problème avec une combinaison linéaire de deux fonctions satisfaisant les conditions aux limites, par exemple :

$$w(x) = a_1 x(1-x) + a_2 \sin\frac{\pi x}{l} \tag{2.23}$$

La formulation résidus pondérés s'écrit cette fois :

$$\forall v \in \mathrm{V} \qquad \int_0^l v(x)\mathrm{ES}\left(2a_1 + a_2\frac{x^2}{l^2}\sin\frac{\pi x}{l} - f\right)\mathrm{d}x = 0 \tag{2.24}$$

Pour pouvoir déterminer les constantes inconnues a_1 et a_2, il nous faut utiliser (au moins) deux fonctions test v. Choisissons deux fonctions linéairement indépendantes, par exemple une constante (comme dans l'exemple 1) et une fonction linéaire :

$$v_1(x) = 1 \qquad v_2(x) = x \tag{2.25}$$

On obtient alors :

$$\begin{aligned}
&\int_0^l \mathrm{ES}\left(2a_1 + a_2\frac{x^2}{l^2}\sin\frac{\pi x}{l} - f\right)\mathrm{d}x = 0 \\
&\int_0^l \mathrm{ES}\left(2xa_1 + a_2\frac{x^3}{l^2}\sin\frac{\pi x}{l} - fx\right)\mathrm{d}x = 0
\end{aligned} \tag{2.26}$$

qui conduit à un système matriciel de deux équations à deux inconnues a_1 et a_2. Une fois résolu (si la solution existe), l'approximation de la solution en déplacements est donnée par le choix de $w(x)$. On peut également choisir deux points de collocation, le premier au tiers et le second aux deux tiers de la barre, c'est-à-dire :

$$v_1(x) = \delta(l/3) \qquad v_2(x) = \delta(2l/3) \tag{2.27}$$

Dans ce cas, on évite les intégrations et le système s'écrit :

$$\begin{aligned}
&\mathrm{ES}\left(2a_1 + a_2\frac{\pi^2}{l^2}\frac{\sqrt{3}}{2}\right) = f\left(\frac{l}{3}\right) \\
&\mathrm{ES}\left(\frac{2l}{3}a_1 + a_2\frac{\pi^2}{l^2}\frac{\sqrt{3}}{6}\right) = f\left(\frac{2l}{3}\right)
\end{aligned} \tag{2.28}$$

Mais on s'aperçoit immédiatement cette fois que ce système ne possède pas de solution (ou une infinité de solutions si le chargement est symétrique), car le déterminant est nul. La méthode de collocation peut donc conduire à une impossibilité.

Reprenons toujours le même exemple en choisissant comme fonctions test celles qui ont servi à l'approximation :

$$v_1(x) = x(l-x) \qquad v_2(x) = \sin\frac{\pi x}{l} \tag{2.29}$$

Dans ce cas, la formulation résidus pondérés conduit à une équation matricielle symétrique.

2.2.2 Poutre console soumise à une charge linéique

Considérons à nouveau une charge linéique f quelconque. Le problème réel s'écrit :

$$\begin{aligned}
&\mathrm{EI}u_{,xxxx} = f \\
&u(0) = 0 \quad u_{,x}(0) = 0 \\
&\mathrm{EI}u_{,xx}(l) = 0 \quad -\mathrm{EI}u_{,xxx}(l) = 0
\end{aligned} \tag{2.30}$$

et on peut encore en trouver la solution exacte u quel que soit le chargement f car on sait résoudre cette équation différentielle du quatrième ordre quelque soit le chargement. Recherchons une solution approchée par résidus pondérés. Choisissons une approximation w du problème réel qui ne satisfasse une fois de plus que les conditions aux limites, c'est-à-dire :

$$w(0)=0; \qquad w_{,x}(0)=0; \qquad \mathrm{EI}w_{,xx}(l)=0; \qquad -\mathrm{EI}w_{,xxx}(l)=0 \tag{2.31}$$

Recherchons cette approximation sur une base polynomiale (au moins quatre fois dérivable) en l'écrivant sous la forme :

$$w(x)=ax^4+bx^3+cx^2+dx+e \tag{2.32}$$

On obtient :

$$w(x)=ax^2(x^2-4lx+6l^2) \tag{2.33}$$

La constante arbitraire a va encore être calculée par la méthode. Le résidu s'écrit :

$$\mathscr{R}(w)=\mathrm{EI}w_{,xxxx}-f=24\mathrm{ES}a-f \tag{2.34}$$

et la formulation résidus pondérés s'écrit, avec la définition de la forme bilinéaire 1 :

$$\forall v\in \mathrm{V} \quad \int_0^l v(x)\big(24\mathrm{EI}a-f\big)\mathrm{d}x=0 \tag{2.35}$$

Comme dans le cas précédent, il reste à choisir une fonction test v. La technique est la même. Si on prend une fonction constante, alors la constante a vaut :

$$a=\frac{1}{24\mathrm{EI}l}\int_0^l f\,\mathrm{d}x \tag{2.36}$$

On obtient la bonne solution lorsque le chargement f est constant ; dans ce cas, la flèche de la poutre est donnée par :

$$w(x)=\frac{fx^2}{24\mathrm{EI}l}(x^2-4lx+6l^2) \tag{2.37}$$

Si le chargement n'est pas constant, cette solution est celle d'une poutre « équivalente » ayant pour chargement la valeur moyenne de f. On pourrait bien sûr choisir d'autres fonctions test et ainsi obtenir d'autres solutions approchées de ce type de problème.

2.2.3 Problèmes de potentiel ou de thermique en deux dimensions

On veut trouver une solution approchée d'un problème de laplacien :

$$\begin{array}{lll} -\Delta u=f & \text{sur} & \Omega \\ u=0 & \text{sur} & \partial\Omega \end{array} \tag{2.38}$$

où $\Omega=\{(x,y)\in\,]-1,\,1[^2\}$. Il nous faut tout d'abord choisir une fonction $w(x,y)$ satisfaisant ces conditions aux limites ; par exemple (a est une constante arbitraire) :

$$w(x,y)=a(x^2-1)(y^2-1) \tag{2.39}$$

Avec ce choix de w, la formulation résidus pondérés s'écrit :

$$\forall v \in \mathrm{V} \qquad \int_\Omega v(x,y)\big(-2a(x^2+y^2-2)-f(x,y)\big)\mathrm{d}\Omega = 0 \tag{2.40}$$

Si on choisit $v(x,y)=1$, la constante a est reliée à la valeur moyenne de la charge $f(x,y)$ sur le domaine et à deux de ses co-moments :

$$\int_\Omega f(x,y)\mathrm{d}\Omega; \quad \int_\Omega x^2 f(x,y)\mathrm{d}\Omega; \quad \int_\Omega y^2 f(x,y)\mathrm{d}\Omega \tag{2.41}$$

Dans le cas particulier où le chargement f est constant, avec ce choix de fonction test constante sur le domaine, on obtient une expression simple de la constante a qui représente le potentiel à l'origine :

$$a = \frac{3f}{10} \tag{2.42}$$

Si on choisit $v(x) = \delta(0)$ (distribution de Dirac au point 0 de coordonnées (0,0)), alors la constante a est relié à la valeur de la charge à l'origine f(0,0) et si le chargement est constant :

$$a = \frac{f}{4} \tag{2.43}$$

ce qui représente une valeur voisine de la solution précédente.

On pourrait choisir d'autres approximations w satisfaisant les conditions aux limites ; par exemple :

$$w(x,y) = a(1+\cos\pi x)(1+\cos\pi y) \tag{2.44}$$

et recommencer les calculs, ou encore utiliser une combinaison linéaire des deux fonctions précédentes, pour obtenir une « meilleure » approximation de la solution du problème réel.

Suivant le choix des fonctions test v, la méthode des résidus pondérés porte plusieurs dénominations rappelées sur la figure 2.1.

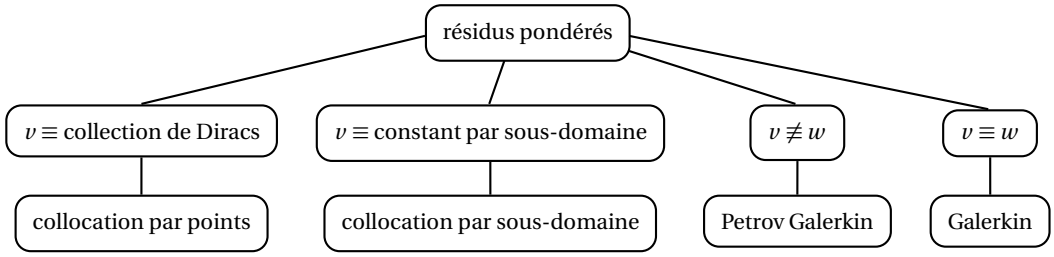


figure 2.1 - *Différentes variations de résidus pondérés*

3

Méthode des éléments finis de type déplacements

3.1 Construction de l'élément fini

La méthode des éléments finis est une méthode de Galerkin avec intégration(s) par parties (IPP). Or l'IPP ne fait que traduire le principe des travaux virtuels. On part donc de :

$$\forall v \in \mathrm{V} \quad \langle v, \mathscr{R}(w) \rangle_1 = 0 \tag{3.1}$$

et après intégration par parties on obtient :

$$\forall v \in \mathrm{V} \quad \langle v, \varphi \rangle_1 = \langle g(v), \sigma \rangle_2 \tag{3.2}$$

qu'on écrit pour quelques v (et non pas pour tout) construits sur la même base que w et on ne satisfait les conditions aux limites qu'après avoir écrit l'équilibre ; il n'y a donc pas a priori de condition du type $\mathscr{B}(w) = t$, ce qui permet un choix plus aisé pour w. Reprenons les quatre premiers exemples du paragraphe 1.3.

3.1.1 Élément fini barre

Le problème réel s'écrit sans préciser les conditions aux limites :

$$-\mathrm{ES}u_{,xx} = f \tag{3.3}$$

et la formulation résidus pondérés est :

$$\forall v \in \mathrm{V} \qquad \int_0^l v(x)\big(-\mathrm{ES}w(x)_{,xx} - f(x)\big)\,\mathrm{d}x = 0 \tag{3.4}$$

soit, après intégration par parties :

$$\forall v \in \mathrm{V} \qquad \int_0^l \mathrm{ES}v(x)_{,x}\, w(x)_{,x}\,\mathrm{d}x - \mathrm{ES}v(x)w(x)_{,x}\,\Big|_0^l - \int_0^l v(x)f(x)\mathrm{d}x = 0 \tag{3.5}$$

Construisons v et w sur la même base de fonctions :

$$v(x) = \mathrm{N}(x)\cdot \mathrm{P} \qquad w(x) = \mathrm{N}(x)\cdot \mathrm{X} \qquad \mathrm{N}(x) = \begin{bmatrix} 1 - \frac{x}{l} & \frac{x}{l} \end{bmatrix} \tag{3.6}$$

où le vecteur des variables nodales X contient les déplacements aux extrémités de l'élément. Tous calculs faits, l'équilibre s'écrit :

$$\mathrm{KX} = \mathrm{F} + \Phi \tag{3.7}$$

où la matrice raideur et les vecteurs forces généralisées sont donnés par :

$$\mathrm{K} = \frac{\mathrm{ES}}{l}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \qquad \mathrm{F} = \begin{pmatrix} \mathrm{F}_1 \\ \mathrm{F}_2 \end{pmatrix} \qquad \Phi = \int_0^l \mathrm{N}^{\mathrm{T}}(x) f(x) \mathrm{d}x \tag{3.8}$$

On reconnaît la matrice raideur classique de l'élément barre ; le vecteur F représente les efforts effectivement appliqués aux nœuds, et le vecteur Φ est le vecteur des charges associées au chargement réparti $f(x)$.

3.1.2 Élément fini poutre

Le problème réel s'écrit sans préciser les conditions aux limites :

$$\mathrm{EI}u_{,xxxx} = f \tag{3.9}$$

et la formulation résidus pondérés est :

$$\forall v \in \mathrm{V} \qquad \int_0^l v(x)\big(\mathrm{EI}w(x)_{,xxxx} - f(x)\big)\mathrm{d}x = 0 \tag{3.10}$$

soit, après deux intégrations par parties :

$$\begin{aligned} \forall v \in \mathrm{V} \qquad \int_0^l \mathrm{EI}v(x)_{,xx}\, w(x)_{,xx}\, \mathrm{d}x + \mathrm{EI}v(x)w(x)_{,xxx}\,\Big|_0^l \\ - \mathrm{EI}v(x)_{,x}\, w(x)_{,xx}\,\Big|_0^l - \int_0^l v(x) f(x) \mathrm{d}x = 0 \end{aligned} \tag{3.11}$$

Construisons v et w sur la même base de fonctions :

$$\begin{aligned} &v(x) = \mathrm{N}(x)\cdot \mathrm{P} \\ &w(x) = \mathrm{N}(x)\cdot \mathrm{X} \\ &\mathrm{N}(x) = \left[1 - \tfrac{3x^2}{l^2} + \tfrac{2x^3}{l^3} \quad x - \tfrac{2x^2}{l} + \tfrac{x^3}{l^2} \quad \tfrac{3x^2}{l^2} - \tfrac{2x^3}{l^3} \quad -\tfrac{x^2}{l} + \tfrac{x^3}{l^2}\right] \end{aligned} \tag{3.12}$$

où le vecteur des variables nodales X contient cette fois les flèches et les rotations aux extrémités de l'élément. Tous calculs faits, l'équilibre s'écrit encore :

$$\mathrm{KX} = \mathrm{F} + \Phi \tag{3.13}$$

où la matrice raideur et les vecteurs forces généralisées sont donnés par :

$$\mathrm{K} = \frac{\mathrm{EI}}{l^3}\begin{bmatrix} 12 & 6l & -12 & 6l \\ 6l & 4l^2 & -6l & 2l^2 \\ -12 & -6l & 12 & -6l \\ 6l & 2l^2 & -6l & 4l^2 \end{bmatrix} \qquad \mathrm{F} = \begin{pmatrix} \mathrm{F}_1 \\ \Gamma_1 \\ \mathrm{F}_2 \\ \Gamma_2 \end{pmatrix} \qquad \Phi = \int_0^l \mathrm{N}^{\mathrm{T}}(x) f(x) \mathrm{d}x \tag{3.14}$$

Le vecteur F représente les efforts et couples effectivement appliqués aux nœuds, et le vecteur Φ est le vecteur des charges associées au chargement réparti $f(x)$.

3.1.3 Élément fini pour laplacien en deux dimensions

Considérons maintenant un élément fini pour les problèmes de laplacien en deux dimensions. Le problème réel s'écrit sans préciser les conditions aux limites :

$$-\Delta u = f \tag{3.15}$$

et la formulation résidus pondérés est :

$$\forall v \in \mathrm{V} \qquad \int_\Omega v(-\Delta w - f)\mathrm{d}\Omega = 0 \tag{3.16}$$

L'intégration par parties en deux dimensions se fait à l'aide de la formule de Green : si a est une fonction scalaire et si b est une fonction vectorielle, on a :

$$\mathrm{div}\, a\, b = a\, \mathrm{div}\, b + \mathrm{grad}\, a \cdot b \tag{3.17}$$

soit, en utilisant de plus le théorème d'Ostrogradski :

$$\forall v \in \mathrm{V} \qquad \int_\Omega \mathrm{grad}\, v\, \mathrm{grad}\, w\, \mathrm{d}\Omega - \int_{\partial\Omega} v\, w_{,n}\, \mathrm{dS} - \int_\Omega v f\, \mathrm{d}\Omega = 0 \tag{3.18}$$

Construisons toujours v et w sur la même base de fonctions

$$u(x,y) = \mathrm{N}(x,y)\cdot \mathrm{P} \qquad w(x,y) = \mathrm{N}(x,y)\cdot \mathrm{X} \tag{3.19}$$

Le choix de l'interpolation dépend de la forme du domaine ; nous y reviendrons dans quelques lignes. La formulation variationnelle devient :

$$\forall \mathrm{P}^\mathrm{T} \qquad \int_\Omega \mathrm{P}^\mathrm{T}\, \mathrm{grad}\, \mathrm{N}^\mathrm{T}\, \mathrm{grad}\, \mathrm{N}\mathrm{X}\mathrm{d}\Omega - \int_{\partial\Omega} \mathrm{P}^\mathrm{T}\mathrm{N}^\mathrm{T} w_{,n}\, \mathrm{dS} - \int_\Omega \mathrm{P}^\mathrm{T}\mathrm{N}^\mathrm{T} f\, \mathrm{d}\Omega = 0 \tag{3.20}$$

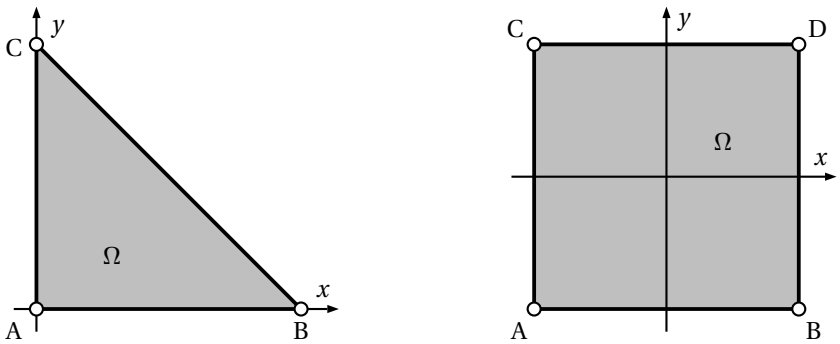
et elle conduit une nouvelle fois à un équilibre de la même forme :

$$\mathrm{KX} = \mathrm{F} + \Phi \tag{3.21}$$

où la matrice raideur est donnée par :

$$\mathrm{K} = \int_\Omega \mathrm{grad}\, \mathrm{N}^\mathrm{T}\, \mathrm{grad}\, \mathrm{N}\mathrm{d}\Omega \tag{3.22}$$

L'expression de cette matrice dépend de l'élément fini choisi. Par exemple choisissons un élément fini T3 de côtés unité illustré sur la figure 3.1(a). Le vecteur des inconnues



(a) triangle isocèle à trois nœuds (b) quadrangle à quatre nœuds

figure 3.1 - *Types d'éléments finis*

nodales contient les valeurs de l'approximation $w(x,y)$ aux nœuds A, B et D :

$$\mathrm{X}^\mathrm{T} = \begin{bmatrix} w(0,0) & w(1,0) & w(0,1) \end{bmatrix} \tag{3.23}$$

L'interpolation est linéaire entre nœuds :

$$\mathrm{N}(x,y) = \begin{bmatrix} 1-x-y & x & y \end{bmatrix} \tag{3.24}$$

On en déduit la matrice raideur élémentaire :

$$\mathrm{K} = \frac{1}{2} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \tag{3.25}$$

Soit un élément fini Q4 défini sur un domaine carré de côtés égaux à deux unités comme montré sur la figure 3.1(b).

Le vecteur des inconnues nodales contient les valeurs de l'approximation w(x,y) aux nœuds A, B, C et D :

$$\mathrm{X}^{\mathrm{T}} = \begin{bmatrix} w(-1,-1) & w(1,-1) & w(1,1) & w(-1,1) \end{bmatrix} \tag{3.26}$$

L'interpolation est construite sur les fonctions de base 1, x, y et xy :

$$\mathrm{N}(x)^{\mathrm{T}} = \frac{1}{4} \begin{bmatrix} (1-x)(1-y) & (1+x)(1-y) & (1+x)(1+y) & (1-x)(1+y) \end{bmatrix} \tag{3.27}$$

Le principe du calcul est le même que pour l'élément triangulaire avec la différence qu'il n'est plus possible d'effectuer l'intégration analytique mais qu'il faut faire appel à l'intégration numérique pour calculer la matrice raideur et le vecteur forces généralisées.

3.1.4 Élastostatique tridimensionnelle

Soit un élément fini pour les problèmes d'élastostatique tridimensionnelle. Le problème réel s'écrit sans préciser les conditions aux limites :

$$-\Delta^* u = f \tag{3.28}$$

Rappelons l'opérateur de Navier :

$$-\Delta^* u = -\operatorname{div}(c_{ijkl} \operatorname{grad}_s u) \tag{3.29}$$

Cette fois, u est un vecteur, grad_s est la partie symétrique du tenseur gradient de ce vecteur, et div est l'opérateur divergence d'un tenseur. La formulation résidus pondérés s'écrit :

$$\forall v \in \mathrm{V} \qquad \int_\Omega v(-\Delta^* w - f)\mathrm{d}\Omega = 0 \tag{3.30}$$

L'intégration par parties se fait toujours à l'aide de la formule de Green ; si a est une fonction vectorielle et si b est une fonction tensorielle, on a :

$$\operatorname{div} a \cdot b = \operatorname{div} b \cdot a + b : \operatorname{grad}^{\mathrm{T}} a$$

où le point $(\cdot)$ représente la simple contraction entre un tenseur et un vecteur et les deux points $(:)$ la double contraction entre deux tenseurs. Identifions a au vecteur u et b au tenseur des contraintes σ défini par :

$$\sigma = c_{ijkl} \operatorname{grad}_s w \tag{3.31}$$

Avec cette notation, on obtient :

$$\forall v \in \mathrm{V} \quad \int_\Omega \mathrm{grad}^{\mathrm{T}} v : \sigma \mathrm{d}\Omega - \int_\Omega \mathrm{div}\, v \cdot \sigma \mathrm{d}\Omega = \int_\Omega v \cdot f \mathrm{d}\Omega \tag{3.32}$$

Décomposons le tenseur grad v en parties symétrique et antisymétrique :

$$\varepsilon = \mathrm{grad}_s\, v \qquad \omega = \mathrm{grad}_a\, v \tag{3.33}$$

En utilisant le fait que le produit doublement contracté entre le tenseur symétrique σ et le tenseur antisymétrique ω est nul, et après utilisation du théorème d'Ostrogradski, on obtient :

$$\forall v \in \mathrm{V} \quad \int_\Omega \varepsilon : \sigma \mathrm{d}\Omega - \int_{\partial\Omega} v \cdot \sigma \cdot n \mathrm{dS} = \int_\Omega v \cdot f \mathrm{d}\Omega \tag{3.34}$$

Construisons toujours v et w sur la même base de fonctions :

$$\begin{aligned} v(x,y,z) &= \mathrm{N}(x,y,z) \cdot \mathrm{P} \\ w(x,y,z) &= \mathrm{N}(x,y,z) \cdot \mathrm{X} \end{aligned} \tag{3.35}$$

Utilisons les notations vectorielles (au lieu de tensorielles) pour les déformations et les contraintes :

$$\begin{aligned} \varepsilon^{\mathrm{T}} &= \begin{bmatrix} \varepsilon_{11} & \varepsilon_{22} & \varepsilon_{33} & \varepsilon_{12} & \varepsilon_{23} & \varepsilon_{31} \end{bmatrix} \\ \sigma^{\mathrm{T}} &= \begin{bmatrix} \sigma_{11} & \sigma_{22} & \sigma_{33} & \sigma_{12} & \sigma_{23} & \sigma_{31} \end{bmatrix} \end{aligned} \tag{3.36}$$

et introduisons la matrice déformation B et la matrice élasticité D telles que :

$$\mathrm{B} = \mathrm{grad}_s \mathrm{N} \qquad \sigma = \mathrm{D} \cdot \varepsilon \tag{3.37}$$

Avec ces notations, on remarque que :

$$\varepsilon^{\mathrm{T}} = \mathrm{P}^{\mathrm{T}} \mathrm{B}^{\mathrm{T}} \qquad \sigma = \mathrm{DBX} \tag{3.38}$$

La formulation variationnelle s'écrit donc :

$$\forall \mathrm{P}^{\mathrm{T}} \quad \int_\Omega \mathrm{P}^{\mathrm{T}} \mathrm{B}^{\mathrm{T}} \mathrm{DBX} \mathrm{d}\Omega - \int_{\partial\Omega} \mathrm{P}^{\mathrm{T}} \mathrm{N}^{\mathrm{T}} \sigma \cdot n \mathrm{dS} = \int_\Omega \mathrm{P}^{\mathrm{T}} \mathrm{N}^{\mathrm{T}} f \mathrm{d}\Omega \tag{3.39}$$

et elle conduit une nouvelle fois à une équation d'équilibre de la même forme :

$$\mathrm{KX} = \mathrm{F} + \Phi \tag{3.40}$$

où la matrice raideur est donnée par :

$$\mathrm{K} = \int_\Omega \mathrm{B}^{\mathrm{T}} \mathrm{DB} \mathrm{d}\Omega \tag{3.41}$$

Cette fois encore nous ne détaillerons pas les calculs : ils sont effectués numériquement et dépendent de l'élément fini et de l'interpolation.

3.2 Quelques résultats théoriques complémentaires

La méthode des éléments finis de type déplacements consiste à choisir arbitrairement une approximation des déplacements que l'on écrit sous la forme :

$$w(x) = \mathrm{N}(x)\mathrm{X} \tag{3.42}$$

Ceci revient en fait à choisir un sous-espace vectoriel V_1 de V, une interpolation $\mathrm{N}(x)$ et un vecteur de variables nodales X. Résumons cette opération par le nouveau schéma à quatre espaces suivant :

$$\begin{array}{ccc} \mathrm{V} & \langle\bullet,\bullet\rangle_1 & \Phi \\ \mathrm{N}\uparrow & & \downarrow \mathrm{N}^{\mathrm{T}} \\ \mathrm{V}_1 & \langle\bullet,\bullet\rangle_1 & \Phi_1 \end{array}$$

où V_1 et Φ_1 représentent les espaces de dimension finie de l'élément « fini » qui représente le domaine continu. Il est ensuite aisé d'en déduire la matrice des déformations notée B telle que $\varepsilon = \mathrm{BX}$; pour chacun des cinq exemples choisis pour illustrer la structure algébrique, on a :

$$\begin{aligned} &\mathrm{B} = \mathrm{N}_{,x} \\ &\mathrm{B} = \mathrm{N}_{,xx} \\ &\mathrm{B} = \mathrm{grad}\,\mathrm{N} \\ &\mathrm{B} = \mathrm{grad}_s\,\mathrm{N} \\ &\mathrm{B} = \mathrm{grad}\,\mathrm{grad}\,\mathrm{N} \end{aligned}$$

que l'on peut résumer par :

$$\mathrm{B} = g \circ \mathrm{N} \tag{3.43}$$

Ceci signifie que dans chaque cas, on écrit la restriction du principe des travaux virtuels à un sous-espace (on l'espère !) vectoriel V_1 de l'espace vectoriel des déplacements V :

$$\forall v \in \mathrm{V}_1 \qquad \langle \mathrm{N}(v), \varphi \rangle_1 = \langle g \circ \mathrm{N}(v), \sigma \rangle_2 \tag{3.44}$$

Et on peut résumer la construction d'un élément fini de type déplacement à l'aide d'un nouveau schéma à quatre espaces :

$$\begin{array}{ccc} & \overset{\mathrm{D}}{\longrightarrow} & \\ \mathrm{E} & \langle\bullet,\bullet\rangle_2 & \Sigma \\ \mathrm{B}\uparrow & & \downarrow \mathrm{B}^{\mathrm{T}} \\ \mathrm{V}_1 & \langle\bullet,\bullet\rangle_1 & \Phi_1 \end{array}$$

Le problème continu est ensuite remplacé par le problème discret suivant :

$$\mathrm{KX} = \mathrm{F} + \Phi \tag{3.45}$$

où la matrice raideur est donnée par :

$$\mathrm{K} = \int_\Omega \mathrm{B}^{\mathrm{T}}\mathrm{DB}\mathrm{d}\Omega \tag{3.46}$$

et l'opérateur D est un scalaire pour les barres ou les poutres (rigidité à la traction ou à la flexion), la matrice identité pour les problèmes de potentiel, et la matrice élasticité (ou celle des plaques) pour des problèmes en deux ou trois dimensions.

Dans le cas où on sait construire les fonctions de Green entre nœuds du problème, on obtient des informations nodales exactes [Wielgosz 82] pour les problèmes d'équilibre ou de dynamique de barres et de poutres. En général, on ne sait pas construire ces fonctions de Green, et on obtient une approximation.

La méthode des éléments finis de type déplacements est donc une approximation de la solution théorique en déformations du problème de l'équilibre :

$$\varepsilon \in \mathrm{I} \cap (\sigma^* + \mathrm{J}) \tag{3.47}$$

Tout ce qui vient d'être fait en déplacements peut être repris en forces : on part dans ce cas d'un sev de dimension finie de l'espace des charges Φ, on interpole le champ des contraintes, et on construit la méthode des éléments finis de type forces qui est une approximation de la solution théorique en contraintes du problème de l'équilibre :

$$\sigma \in d(\mathrm{I}) \cap (\sigma^* + \mathrm{J}) \tag{3.48}$$

On remarquera que dans les deux cas, on travaille au niveau des espaces E et Φ.

3.3 Rappels sur l'assemblage des éléments finis

Le lecteur trouvera de nombreux exemples dans [Wielgosz 99]. Donnons ici trois exemples simples d'assemblage (barres, poutres et problèmes de potentiel).

3.3.1 Barreau bi-encastré soumis à une charge linéique constante

Il nous faut utiliser au moins deux éléments finis ; choisissons les de même longueur. On note u_i $(i = 1, \ldots, 3)$ les déplacements des nœuds un à trois. L'équilibre d'un élément de longueur l s'écrit :

$$\frac{\mathrm{ES}}{l} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{pmatrix} u_i \\ u_{i+1} \end{pmatrix} = \begin{pmatrix} \mathrm{F}_i \\ \mathrm{F}_{i+1} \end{pmatrix} + \frac{fl}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{3.49}$$

où F_i et F_{i+1} représentent les efforts appliqués aux nœuds de l'élément i. Après assemblage [1], on obtient :

$$\frac{2\mathrm{ES}}{l} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ u_2 \\ 0 \end{pmatrix} = \begin{pmatrix} \mathrm{F}_1 \\ 0 \\ \mathrm{F}_3 \end{pmatrix} + \frac{fl}{4} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \tag{3.50}$$

Le déplacement u_2 en $x = l/2$ [2] est donné par la deuxième ligne :

$$u_2 = \frac{fl^2}{8\mathrm{ES}} \tag{3.51}$$

et les efforts aux nœuds 1 et 3 (exacts) sont obtenus par les deux autres lignes :

$$\mathrm{F}_1 = \mathrm{F}_3 = \frac{fl}{2} \tag{3.52}$$

1. L'assemblage consiste à écrire les équations de liaison entre les éléments et éliminer des efforts correspondants.
2. Nous rappelons que dans ce cas précis, le déplacement calculé est exact.

3.3.2 Poutre bi-encastrée soumise à une charge linéique constante

Afin de simplifier les notations, considérons une poutre de longueur $\mathrm{L} = 2l$. Choisissons un chargement linéique d'intensité constante p. Discrétisons la poutre en deux éléments de longueur l. Déterminons tout d'abord le vecteur forces généralisées associé au chargement réparti d'intensité constante p. Rappelons qu'il est donné, pour un élément de longueur l par :

$$\Phi = \int_0^l \mathrm{N}^{\mathrm{T}} p \,\mathrm{d}x \tag{3.53}$$

ce qui donne, lorsque p est constant :

$$\Phi^{\mathrm{T}} = \begin{bmatrix} \frac{pl}{2} & \frac{pl^2}{12} & \frac{pl}{2} & -\frac{pl^2}{12} \end{bmatrix} \tag{3.54}$$

L'assemblage se fait de la même manière que pour le barreau mais en travaillant avec deux degrés de liberté par nœuds au lieu d'un seul. Écrivons l'équilibre des deux éléments assemblés :

$$\frac{\mathrm{EI}}{l^3} \begin{bmatrix} 12 & 6l & -12 & 6l & 0 & 0 \\ 6l & 4l^2 & -6l & 2l^2 & 0 & 0 \\ -12 & -6l & 24 & 0 & -12 & 6l \\ 6l & 2l^2 & 0 & 8l^2 & -6l & 2l^2 \\ 0 & 0 & -12 & -6l & 12 & -6l \\ 0 & 0 & 6l & 2l^2 & -6l & 4l^2 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ v_2 \\ \theta_2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \mathrm{R}_1 \\ \Gamma_1 \\ 0 \\ 0 \\ \mathrm{R}_3 \\ \Gamma_3 \end{pmatrix} + \begin{pmatrix} \frac{pl}{2} \\ \frac{pl^2}{12} \\ pl \\ 0 \\ \frac{pl}{2} \\ -\frac{pl^2}{12} \end{pmatrix} \tag{3.55}$$

Le système est résolu par la méthode des sous-matrices ; les équations 3 et 4 donnent les valeurs de la flèche et de la rotation au nœud 2 :

$$v_2 = \frac{pl^4}{24\mathrm{EI}}; \qquad \theta_2 = 0 \tag{3.56}$$

Les quatre autres équations du système donnent ensuite les valeurs des réactions :

$$\mathrm{R}_1 = -pl; \qquad \Gamma_1 = -\frac{pl^2}{3}; \qquad \mathrm{R}_3 = -pl; \qquad \Gamma_3 = \frac{pl^2}{3} \tag{3.57}$$

3.3.3 Problème de thermique stationnaire

Le problème est défini en deux dimensions sur un domaine carré plan Ω délimité par les droites d'équation $x = \pm 1$ et $y = \pm 1$ comme décrit sur la figure 3.1(b). Les sources de chaleur internes f sont constantes et on suppose des conditions aux limites de Dirichlet.

On discrétise le domaine en quatre triangles rectangles isocèles de côtés égaux à $\sqrt{2}$ à l'image de la figure 3.1(a). Chaque domaine triangulaire est dénommé T. L'interpolation est ici :

$$\mathrm{N}(x,y)^{\mathrm{T}} = \begin{bmatrix} 1 - \frac{x}{\sqrt{2}} - \frac{y}{\sqrt{2}} & \frac{x}{\sqrt{2}} & \frac{y}{\sqrt{2}} \end{bmatrix} \tag{3.58}$$

Mais la matrice raideur élémentaire est identique à celle établie au paragraphe 3.1 :

$$\mathrm{K} = \frac{1}{2}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \tag{3.59}$$

Le vecteur force généralisée associé au chargement f est :

$$\Phi = \int_{\mathrm{T}} \mathrm{N}^{\mathrm{T}} f \,\mathrm{d}x = \frac{f}{3}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{3.60}$$

La matrice raideur est invariante par rotation autour d'un axe perpendiculaire au plan du domaine ; elle peut donc être utilisée pour chacun des triangles. Après assemblage des quatre matrices élémentaires, des vecteurs forces généralisées, on écrit l'équilibre du domaine en tenant compte des conditions aux limites :

$$\begin{bmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} w_1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \mathrm{F}_2 \\ \mathrm{F}_3 \\ \mathrm{F}_4 \\ \mathrm{F}_5 \end{pmatrix} + \frac{f}{3}\begin{pmatrix} 4 \\ 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} \tag{3.61}$$

La valeur du champ de température à l'origine est $w_1 = f/3$. On remarquera que ce résultat est très voisin de ceux que nous avions obtenus par résidus pondérés.

Deuxième partie

ÉQUATIONS INTÉGRALES

Bernard Peseux

La modélisation d'un grand nombre de situations d'intérêt pratique pour l'ingénieur ou chercheur, conduit à la recherche de solutions d'équations aux dérivées partielles, assorties de conditions aux limites et de conditions initiales, notamment en *Mécanique du Solide*, en *Mécanique des fluides*, en *Acoustique*, en *Thermique* ou en *Électromagnétisme*.

Ces équations sont posées en général sur des domaines géométriques, qui ne permettent pas l'emploi des techniques classiques de recherche de solutions exactes et elles doivent être résolues par des méthodes numériques et en particulier par la méthode des éléments finis, des équations intégrales (éléments de frontière, méthode des singularités) ou des différences finies.

L'étude des équations intégrales a commencé il y a plus d'un siècle (Théorie du potentiel, identité de Somigliana), mais les développements concernant la résolution numérique ne datent que des années 1960.

4

Principe et champs d'application

4.1 Principe

Un problème aux limites, d'inconnues u, posé sur un domaine $\Omega \subset \mathrm{R}^n$, $(n = 1,2,3)$ et associé à un opérateur aux dérivées partielles d'ordre pair $(2m)$, auto-adjoint $\mathscr{L}$, présente typiquement, comme cela a été vu dans la partie I la structure suivante :

$$\begin{array}{lll} \mathscr{L}(u)+f=0 & \text{dans} & \Omega \\ u=u_d & \text{sur} & \partial\Omega_u \\ \mathscr{B}(u)+q=0 & \text{sur} & \partial\Omega_\sigma \end{array} \tag{4.1}$$

avec $\partial\Omega_u \cup \partial\Omega_\sigma = \partial\Omega$ et $\partial\Omega_u \cap \partial\Omega_\sigma = \varnothing$ et où $\mathscr{B}$ est un opérateur différentiel (ou système d'opérateurs) d'ordre au plus $2m-1$.

Par exemple dans le cas de l'équation de Poisson, $\mathscr{L}$ est l'opérateur de Laplace Δ et $\mathscr{B}$, l'opérateur dérivée normale, $\frac{\partial}{\partial n}$, avec $\vec{n}$ normale à $\partial\Omega$, extérieure à Ω. Dans ce cas particulier, soit u une approximation de u_{exact} et v une fonction de pondération. La méthode des résidus pondérés, donne la forme intégrale :

$$\mathrm{W}(v) = \int_\Omega v(\Delta u + f)\mathrm{d}\Omega = \int_\Omega v\Delta u\,\mathrm{d}\Omega + \int_\Omega v f\,\mathrm{d}\Omega \tag{4.2}$$

Une première intégration par parties généralisées (application de la formule de Green) appliquée à la première intégrale donne :

$$\int_\Omega v\Delta u\,\mathrm{d}\Omega + \int_\Omega \vec{\mathrm{grad}}v \cdot \vec{\mathrm{grad}}u\,\mathrm{d}\Omega = \int_{\partial\Omega} v\frac{\partial u}{\partial n}\mathrm{d}\Gamma \tag{4.3}$$

Une seconde intégration par parties sur le second terme donne :

$$\int_\Omega \vec{\mathrm{grad}}v \cdot \vec{\mathrm{grad}}u\,\mathrm{d}\Omega + \int_\Omega u\Delta v\mathrm{d}\Omega = \int_{\partial\Omega} u\frac{\partial v}{\partial n}\mathrm{d}\Gamma \tag{4.4}$$

et en retranchant cette dernière expression (4.4) à la précédente (4.3), on obtient la seconde formule de Green :

$$\int_\Omega (v\Delta u - u\Delta v)\mathrm{d}\Omega = \int_{\partial\Omega} \left(v\frac{\partial u}{\partial n} - u\frac{\partial v}{\partial n}\right)\mathrm{d}\Gamma \tag{4.5}$$

On peut généraliser cette expression au cas des opérateurs $\mathscr{L}$ de l'équation locale et $\mathscr{B}$ des conditions aux limites et on obtient une relation (formule de réciprocité) de la forme :

$$\int_\Omega \Big(v\mathscr{L}(u) - u\mathscr{L}(v)\Big)\mathrm{d}\Omega = \int_{\partial\Omega} \Big(v\mathscr{B}(u) - u\mathscr{B}(v)\Big)\mathrm{d}\Gamma \tag{4.6}$$

Remplaçons v par une solution particulière dite élémentaire qui vérifie l'équation locale, pour une source ponctuelle $f(y)=\delta(y-x)$ appliquée en un point fixé $x \notin \partial\Omega$. Notons $G(x,y)$ cette solution élémentaire, fonction du point courant y. L'équation locale de (4.1) s'écrit :

$$\Delta G(x,y)+\delta(y-x)=0 \tag{4.7}$$

La mesure de Dirac $\delta(y-x)$ est définie par la propriété :

$$\int_\Omega \delta(y-x)u(y)\mathrm{d}V_y = \alpha u(x) \quad \text{avec} \quad \begin{cases} \alpha=1 \text{ si } x\in\Omega \\ \alpha=0 \text{ si } x\notin\Omega \end{cases} \tag{4.8}$$

et la substitution de v par G dans (4.5), conduit à :

$$\alpha u(x)=\int_{\partial\Omega}\left(G(x,y)\frac{\partial u(y)}{\partial n}-u(y)\frac{\partial G(x,y)}{\partial n}\right)\mathrm{d}\Gamma+\int_\Omega G(x,y)f(y)\mathrm{d}\Omega \tag{4.9}$$

Cette formule de représentation intégrale donne explicitement la valeur de u en tout point x intérieur Ω à comme somme :

- d'une intégrale de domaine contenant la source f, solution particulière de l'équation locale de (4.1) ne vérifiant pas en général les conditions aux limites ;
- d'un terme intégral Ω sur $\partial\Omega$ faisant intervenir les valeurs de u et $\frac{\partial u}{\partial n}$ pour moitié prescrites par les conditions aux limites et pour moitié inconnues. Le calcul de u dans est ramené à celui des valeurs de $(u,\frac{\partial u}{\partial n})$ sur $\partial\Omega$ restées inconnues après prise en compte des conditions aux limites. *On a « gagné » une dimension d'espace.*

4.1.1 Équilibre global

Pour toute solution élémentaire G vérifiant (4.7), pour tout domaine Ω de frontière $\partial\Omega$ et pour tout point $x \notin \partial\Omega$, l'intégration sur de l'équation locale (4.7) et l'application de la formule de la divergence conduisent à l'identité :

$$\alpha+\int_{\partial\Omega} H(x,y)\mathrm{d}S_y = 0 \quad \text{avec} \quad \begin{cases} \alpha=1 \text{ si } x\in\Omega \\ \alpha=0 \text{ si } x\notin\Omega \end{cases} \tag{4.10}$$

qui exprime l'équilibre entre le flux H à travers $\partial\Omega$ et la source ponctuelle unitaire exercée en x. Le flux H est donné par :

$$H(x,y)=G_{,j}(x,y)n_j(y)=\frac{\partial G(x,y)}{\partial n} \tag{4.11}$$

4.1.2 Principe de la méthode

La première étape consiste à rechercher des solutions élémentaires. Ensuite, l'équation (4.9) n'est *a priori* valable que pour $x \notin \partial\Omega$. Il reste donc à formuler une équation intégrale de frontière qui ne porterait que sur les valeurs à la frontière de u et $\frac{\partial u}{\partial n}$, mais les fonctions G solutions élémentaires sont telles que $\frac{\partial G}{\partial n}$ présente en $x \in \partial\Omega$ une singularité non intégrable. L'obtention d'une équation intégrale de frontière impose donc un passage à la limite. Ce passage à la limite que nous n'aborderons pas ici conduit à la forme intégrale régularisée qui s'écrit :

$$\int_{\partial\Omega}\left(G(x,y)\frac{\partial u(y)}{\partial n}-[u(y)-u(x)]\frac{\partial G(x,y)}{\partial n}\right)\mathrm{d}\Gamma+\int_\Omega G(x,y)f(y)\mathrm{d}\Omega=0 \tag{4.12}$$

ou bien :

$$\alpha u(x) = \int_{\partial\Omega} \left(\mathrm{G}(x,y) \frac{\partial u(y)}{\partial n} - u(y) \frac{\partial \mathrm{G}(x,y)}{\partial n} \right) \mathrm{d}\Gamma + \int_{\Omega} \mathrm{G}(x,y) f(y) \mathrm{d}\Omega \tag{4.13}$$

avec dans le cas général :

$$\alpha = \lim_{\varepsilon \to 0} \int_{\mathrm{S}_\varepsilon} \frac{\partial \mathrm{G}(x,y)}{\partial n} \mathrm{dS}_y \tag{4.14}$$

qui donne $\alpha = \frac{1}{2}$ lorsque la frontière $\partial\Omega$ est suffisamment régulière, *ie* possède en tout point un plan tangent. L'intégrale (4.12) redonne pour $x \notin \partial\Omega$ la représentation intégrale (4.9), elle couvre donc toutes les positions de x. L'équation intégrale (4.12) (ou bien (4.13)) ne fait intervenir que des intégrales convergentes qui présentent des singularités pour $y = x$, intégrables.

Une fois ce passage à la limite accompli, la méthode des équations intégrales procède en deux temps :

1. résolution de l'équation intégrale de frontière, permettant le calcul du couplet $(u, \frac{\partial u}{\partial n})$ en tout point de $\partial\Omega$;
2. application de la formule de représentation intégrale (4.9) permettant le calcul explicite de la valeur de $u(x)$ en tout point x intérieur à Ω.

4.2 Formulations directes et indirectes

La méthode des équations intégrales peut être formulée avec deux approches différentes : une approche *directe* et une approche *indirecte*.

L'approche directe (méthode des éléments de frontière) consiste à exprimer le champ inconnu en un point intérieur du domaine en fonction de toutes les conditions aux limites (naturelles et cinématiques). Cette formulation rigoureuse, peut découler ou bien de l'application du théorème de Green ou d'une mise en œuvre particulière de la méthode des résidus pondérés.

La méthode indirecte (méthode des singularités) consiste à remplacer la frontière du domaine par une distribution inconnue de singularités, telle que les conditions aux limites sur la frontière soient restituées.

À partir de la formulation directe, il est possible de retrouver la formulation indirecte.

4.3 Domaines d'applications

4.3.1 Conduction de la chaleur

La température T d'un milieu de conductivité k, homogène mais éventuellement anisotrope (k est alors un tenseur d'ordre 2 symétrique), régi par la loi de Fourier :

$$q = -k \vec{\mathrm{grad}} \mathrm{T} \tag{4.15}$$

vérifie l'équation locale :

$$\mathrm{div}(k \vec{\mathrm{grad}} \mathrm{T}) + f = 0 \tag{4.16}$$

qui dans le cas d'une conductivité isotrope se réduit à l'équation de Poisson :

$$k\Delta \mathrm{T} + f = 0 \tag{4.17}$$

En régime transitoire, la température dépend également du temps et vérifie l'équation locale de la diffusion :

$$-\rho c \frac{\partial \mathrm{T}}{\partial t} + \mathrm{div}(k\vec{\mathrm{grad}}\mathrm{T}) + f = 0 \tag{4.18}$$

où f est la source volumique de chaleur, c, la chaleur massique - masse volumique et q, le flux thermique. Les conditions aux limites sont des conditions de flux imposé et de température imposée :

$$\begin{array}{lll} \mathrm{T} = \mathrm{T}_d & \text{sur} & \partial\Omega_u \\ \dfrac{\partial \mathrm{T}}{\partial n} + q_s = 0 & \text{sur} & \partial\Omega_\sigma \end{array} \tag{4.19}$$

4.3.2 Écoulement de fluide

Fluide parfait

Le champ de vitesses $\mathrm{U}(x,t)$ d'un écoulement de fluide parfait incompressible et irrotationnel dérive d'un potentiel (x,t) tel que $\vec{\mathrm{U}} = -\vec{\mathrm{grad}}\Phi$. Ce potentiel des vitesses satisfait l'équation de Laplace :

$$\Delta\Phi = 0 \tag{4.20}$$

Cette équation locale est complétée par des conditions cinématiques aux interfaces fluide-solide Γ_{fs}, qui traduisent les conditions d'imperméabilité des parois solides : égalité des vitesses normales des particules fluides et solides :

$$\frac{\partial \Phi}{\partial n} = \vec{\mathrm{V}}_s \cdot \vec{n} \quad \text{sur} \quad \Gamma_{fs} \tag{4.21}$$

Acoustique

Le potentiel des vitesses (x,t) d'un fluide parfait compressible, avec l'hypothèse de petits mouvements, satisfait l'équation de Helmoltz (équation des ondes) :

$$\Delta\Phi - \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2} = 0 \tag{4.22}$$

c étant la célérité des ondes dans le fluide.

Écoulement souterrain

L'écoulement souterrain d'un fluide parfait incompressible et suivant la loi de Darcy :

$$\vec{\mathrm{U}} = -k\vec{\mathrm{grad}}\Phi \tag{4.23}$$

où $\Phi = \frac{p}{\rho} + gz$ est la charge statique et k, le tenseur des conductivités hydrauliques, est gouverné par l'équation de Laplace :

$$k\Delta\Phi = 0 \quad \text{ou} \quad \mathrm{div}(k\vec{\mathrm{grad}})\Phi = 0 \tag{4.24}$$

Écoulement lent de fluides visqueux incompressibles

Les champs de vitesse U et de pression p d'un fluide incompressible de viscosité vérifient la condition d'incompressibilité et l'équation de Navier-Stockes qui pour les mouvements lents stationnaires et en l'absence de gravité, s'écrivent :

$$\begin{aligned} &\mu\Delta\vec{\mathrm{U}} - \vec{\mathrm{grad}}p = 0 \\ &\mathrm{div}\,\vec{\mathrm{U}} = 0 \end{aligned} \tag{4.25}$$

4.3.3 Électromagnétisme

Les champs électrique $\vec{\mathrm{E}}$ et magnétique $\vec{\mathrm{B}}$ ainsi que la densité de charge électrique ρ dans un milieu linéaire, isotrope, homogène et stationnaire, avec σ la conductivité électrique, vérifient les équations de Maxwell :

$$\begin{aligned} &\mathrm{div}\,\vec{\mathrm{E}} = \frac{\rho}{\varepsilon_0(1+\chi_e)} \qquad && \vec{\mathrm{rot}}\,\vec{\mathrm{E}} + \frac{\partial\vec{\mathrm{B}}}{\partial t} = 0 \\ &\mathrm{div}\,\vec{\mathrm{B}} = 0 && \vec{\mathrm{rot}}\,\vec{\mathrm{B}} - \frac{1}{c^2}\frac{\partial\vec{\mathrm{E}}}{\partial t} = \mu_0(1+\chi_m)\sigma\vec{\mathrm{E}} \end{aligned} \tag{4.26}$$

c est la célérité des ondes magnétiques dans le milieu considérée et ε_0 et μ_0 sont les constantes du vide.

En combinant les équations ci-dessus, les champs $\vec{\mathrm{E}}$ et $\vec{\mathrm{B}}$ vérifient deux équations aux dérivées partielles faisant intervenir l'opérateur de Laplace. Ces équations dans le cas de l'électrostatique, se simplifient. Le champ électrique $\vec{\mathrm{E}}$ dérive d'un potentiel scalaire V qui vérifie l'équation de Poisson :

$$\Delta\mathrm{V} = \frac{\rho}{\varepsilon_0(1+\chi_e)} \tag{4.27}$$

4.3.4 Élastostatique linéaire

L'équilibre élastique d'un solide occupant le domaine avec l'hypothèse des petites perturbations, est caractérisé par les équations :

$$\begin{aligned} &\sigma_{ij,j} + \rho f_i = 0 \quad (\vec{\mathrm{div}}\sigma + \rho\vec{f} = \vec{0}) \\ &\varepsilon_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}) \\ &\sigma_{ij} = \mathrm{C}_{ijkl}\varepsilon_{kl} \end{aligned} \tag{4.28}$$

Tout triplet (u, σ, f) sera qualifié d'état élastostatique sur Ω. Les relations précédentes combinées, conduisent à l'équation de Navier (ou Lamé-Navier) gouvernant les déplacements :

$$\vec{\Delta}^*\vec{u} + \rho\vec{f} = \vec{0} \tag{4.29}$$

avec $\vec{\Delta}^*\vec{u} = \mathrm{C}_{ijkl}\,u_{k,lj}\,\vec{e}_i$. Les conditions aux limites sur la frontière $\partial\Omega$ sont :

$$\begin{aligned} &\vec{u} = \vec{u}_d \quad \text{sur} \quad \partial\Omega_u \\ &\sigma\cdot\vec{n} = \vec{\mathrm{T}}_d \quad \text{sur} \quad \partial\Omega_\sigma \end{aligned} \tag{4.30}$$

En élasticité isotrope, le tenseur d'élasticité $\overline{\overline{C}}$ ne dépend que de deux constantes élastiques et constantes de Lamé ou E et ν (module d'Young et coefficient de Poisson) et l'équation locale précédente donne les équations dites de *Lamé-Navier* :

$$(\lambda+\mu)\left(u_{k,k}\right)_{,i}+\mu u_{i,kk}+\rho f_i=0 \tag{4.31}$$

On pourra aussi utiliser ces équations sous la forme :

$$(\lambda+\mu)\,\vec{\text{grad}}(\text{div}\,\vec{u})+\mu\Delta\vec{u}+\rho\vec{f}=\vec{0} \tag{4.32}$$

ou encore :

$$(\lambda+2\mu)\,\vec{\text{grad}}(\text{div}\,\vec{u})-\mu\vec{\text{rot}}\left(\vec{\text{rot}}\,\vec{u}\right)+\rho\vec{f}=\vec{0} \tag{4.33}$$

Si le champ de déplacement est irrotationnel, les équations se réduisent à :

$$(\lambda+2\mu)\,\vec{\text{grad}}(\text{div}\,\vec{u})+\rho\vec{f}=\vec{0} \tag{4.34}$$

ce qui implique que $\vec{\text{rot}}\,\rho\vec{f}=\vec{0}$ et que $\rho\vec{f}$ est de la forme $\rho\vec{f}=\vec{\text{grad}}\Phi$. Il s'ensuit qu'on obtient après intégration :

$$(\lambda+2\mu)\,\text{div}\,\vec{u}+\Phi=\text{cste} \tag{4.35}$$

5

Équation de poisson

5.1 Équations intégrales

5.1.1 Problème intérieur et fonction harmonique dans un domaine borné

Dans ce cas l'équation intégrale est celle établie en (4.12) :

$$\int_{\partial\Omega}\left(\mathrm{G}(x,y)\frac{\partial u(y)}{\partial n}-\big(u(y)-u(x)\big)\frac{\partial \mathrm{G}(x,y)}{\partial n}\right)\mathrm{d}\Gamma+\int_{\Omega}\mathrm{G}(x,y)f(y)\mathrm{d}\Omega=0 \tag{5.1}$$

ou bien (4.13) qui dans le cas d'une frontière régulière donne :

$$\int_{\partial\Omega}\left(u(y)\frac{\partial \mathrm{G}(x,y)}{\partial n}-\mathrm{G}(x,y)\frac{\partial u(y)}{\partial n}\right)\mathrm{dS}-\int_{\Omega}\mathrm{G}(x,y)f(y)=\begin{cases}-u(x) & \text{si } x\in\Omega_i\\ -\dfrac{u(x)}{2} & \text{si } x\in\partial\Omega\\ 0 & \text{sinon}\end{cases} \tag{5.2}$$

avec $\vec{n}$ normale extérieure au domaine i.

5.1.2 Problème extérieur et fonction harmonique dans un domaine non borné

On considère une sphère de surface Σ entourant complètement le domaine Ω_i, dont le rayon tend vers l'infini. On applique les résultats de l'équation (5.2) au domaine Ω'_e compris entre les surfaces $\partial\Omega$ et Σ avec une normale $\vec{n}$ extérieure au domaine Ω'_e. On obtient alors :

$$\begin{aligned}&\int_{\partial\Omega}\left(u(y)\frac{\partial \mathrm{G}(x,y)}{\partial n}-\mathrm{G}(x,y)\frac{\partial u(y)}{\partial n}\right)\mathrm{dS}\\ &-\int_{\Omega}\mathrm{G}(x,y)f(y)+\int_{\Sigma}\left(u(y)\frac{\partial \mathrm{G}(x,y)}{\partial n}-\mathrm{G}(x,y)\frac{\partial u(y)}{\partial n}\right)\mathrm{dS}\end{aligned}=\begin{cases}-u(x) & \text{si } x\in\Omega'_e\\ -\dfrac{u(x)}{2} & \text{si } x\in\partial\Omega\\ 0 & \text{sinon}\end{cases} \tag{5.3}$$

Si $\mathrm{G}(x,y)$, fonction de Green du problème, satisfait les conditions de décroissance à l'infini alors :

$$\int_{\partial\Omega}\left(u(y)\frac{\partial \mathrm{G}(x,y)}{\partial n}-\mathrm{G}(x,y)\frac{\partial u(y)}{\partial n}\right)\mathrm{dS}-\int_{\Omega}\mathrm{G}(x,y)f(y)=\begin{cases}-u(x) & \text{si } x\in\Omega_i\\ -\dfrac{u(x)}{2} & \text{si } x\in\partial\Omega\\ 0 & \text{sinon}\end{cases} \tag{5.4}$$

avec $\vec{n}$ normale extérieure au domaine Ω'_e, donc intérieur au domaine intérieur Ω_i.

5.2 Solutions élémentaires

En milieu illimité (espace infini), r désignant la distance euclidienne MP soit $|r| = \left|y - x\right|$ et la condition à l'infini s'écrit :

$$\lim_{r\to 0} u = \mathcal{O}\left(\frac{1}{r}\right) \tag{5.5}$$

et la fonction de Green associée est la fonction :

$$\mathrm{G}(x,y) = \frac{1}{4\pi r} \quad \text{ou bien} \quad \mathrm{G}(x,y) = \frac{1}{r} \tag{5.6}$$

le flux correspondant est :

$$\mathrm{H}(x,y) = -\frac{1}{4\pi r^2} r_{,n} \quad \text{ou bien} \quad \mathrm{H}(x,y) = -\frac{1}{r^2} r_{,n} \tag{5.7}$$

Dans le cas du demi-espace $x_3 \leqslant 0$, les solutions élémentaires peuvent être construites par la méthode des images. Soit M le point de coordonnées $x = (x_1, x_2, x_3)$ et M′ le point symétrique de M par rapport au plan $x_3 = 0$, les coordonnées de M′ sont $\bar{x} = (x_1, x_2, -x_3)$. En notant $\bar{r}$ la distance M′P, la fonction de Green vérifiant une condition de Dirichlet [1] est :

$$\mathrm{G}(\mathrm{M},\mathrm{P}) = \frac{1}{r} - \frac{1}{\bar{r}} \tag{5.8}$$

et vérifiant une condition de Neumann [2] est :

$$\mathrm{G}(\mathrm{M},\mathrm{P}) = \frac{1}{r} + \frac{1}{\bar{r}} \tag{5.9}$$

5.3 Formulation directe

Dans un premier temps pour simplifier, on suppose que les distributions volumiques $f(y)$ sont nulles et on cherche le couplet $(u, q = \frac{\partial u}{\partial n})$ solution de :

$$\int_{\partial\Omega} \left(\mathrm{G}(x,y) \frac{\partial u(y)}{\partial n} - \big(u(y) - u(x)\big) \frac{\partial \mathrm{G}(x,y)}{\partial n} \right) \mathrm{d}\Gamma = 0 \tag{5.10}$$

le point x appartenant à la frontière $\partial\Omega$. La méthode mise en œuvre ici étant une méthode de collocation.

5.3.1 Discrétisation géométrique

Par une démarche analogue à celle des éléments finis, la frontière $\partial\Omega$ est discrétisée en éléments de frontière $\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_\mathrm{N}$. Ces éléments de type linéique (en bidimensionnel) ou surfacique (ou tridimensionnel) sont définis à partir d'un élément de référence par une transformation géométrique qui fait appel aux fonctions de forme :

$$x = \sum_{i=1}^{n} \mathrm{M}_i(\xi) x_i = \mathbf{M}(\xi)\mathbf{x} \tag{5.11}$$

L'intégrale (5.10) est alors écrite comme une somme d'intégrales élémentaires calculées sur chaque élément frontière :

$$\sum_{e=1}^{\mathrm{N}} \int_{\mathrm{E}_e} \left(\mathrm{G}(x,y) \frac{\partial u(y)}{\partial n} - \big(u(y) - u(x)\big) \frac{\partial \mathrm{G}(x,y)}{\partial n} \right) \mathrm{d}\Gamma = 0 \tag{5.12}$$

1. $\mathrm{G}(\mathrm{M},\mathrm{P}) = 0$ en $x_3 = 0$
2. $\mathrm{H}(\mathrm{M},\mathrm{P}) = 0$ en $x_3 = 0$

5.3.2 Discrétisation des inconnues

La représentation des variables (u,q) reprend encore le formalisme de la méthode des éléments finis. Sur chaque élément frontière E_e, on réalise l'interpolation de ces variables en fonction de leurs valeurs nodales sur l'élément :

$$\begin{aligned} u &= \sum_i \mathrm{N}_i(\xi) u_{ie} = \mathbf{N}(\xi)\mathbf{u}_e \\ q &= \sum_i \mathrm{N}_i(\xi) q_{ie} = \mathbf{N}(\xi)\mathbf{q}_e \end{aligned} \tag{5.13}$$

Comme pour la méthode des éléments finis, on parlera d'interpolation conforme, non conforme ou isoparamétrique.

5.3.3 Construction du système discrétisé et résolution numérique

Discrétisation de l'équation intégrale

La méthode des collocation consiste à forcer l'équation (5.12) à être vérifiée exactement en certains points : *les points de collocation* de coordonnées x_c qui en général pour les problèmes qui nous intéressent, sont choisis comme étant les nœuds du maillage de coordonnées x_i. En reportant les interpolations dans (5.12), il vient :

$$\sum_{e=1}^{\mathrm{N}} \int_{\mathrm{E}_e} \left(\mathrm{G}(x_c,y) \frac{\partial u(y)}{\partial n} - \big(u(y)-u(x_c)\big) \frac{\partial \mathrm{G}(x_c,y)}{\partial n} \right) \mathrm{d}\Gamma = 0 \tag{5.14}$$

Dans (5.14) [3], il apparaît des intégrales élémentaires régulières lorsque $x_c \notin \mathrm{E}_e$ et des intégrales singulières lorsque $x_c \in \mathrm{E}_e$. Pour un point $x \in \partial\Omega$, posons :

$$\mathrm{I}(x) = \{ e \in [1,\mathrm{N}], x \in \mathrm{E}_e \} \quad \text{et} \quad \bar{\mathrm{I}}(x) = \{ [1,\mathrm{N}] - \mathrm{I}(x) \} \tag{5.15}$$

En séparant les intégrales régulières et les intégrales singulières, l'équation discrétisée précédente s'écrit :

$$\begin{aligned} &\sum_{e\in \mathrm{I}(x)} \sum_{p=1}^{\mathrm{N}} \big(\mathrm{AS}(e,p) u_{i(e,p)} - \mathrm{B}(e,p) q_{i(e,p)} \big) \\ &\qquad + \sum_{e\in \bar{\mathrm{I}}(x)} \sum_{p=1}^{\mathrm{N}} \big(\mathrm{AR}(e,p) u_{i(e,p)} - \mathrm{B}(e,p) q_{i(e,p)} \big) - u(x_c) \sum_{e\in \bar{\mathrm{I}}(x)} \hat{\mathrm{H}}(e) \end{aligned} \tag{5.16}$$

avec :

$$\begin{aligned} \mathrm{AS}(e,p) &= \int_{\mathrm{E}_e} \big(\mathrm{N}_p(\xi) - \mathrm{N}_p(\eta_\varepsilon) \big) \mathrm{H}(x_c,y) \mathrm{dS}_y \quad & \mathrm{B}(e,p) &= \int \mathrm{N}_p(\xi) \mathrm{G}(x_c,y) \mathrm{dS}_y \\ \mathrm{AR}(e,p) &= \int \mathrm{N}_p(\xi) \mathrm{H}(x_c,y) \mathrm{dS}_y & \hat{\mathrm{H}}(e) &= \int \mathrm{H}(x_c,y) \mathrm{dS}_y \end{aligned} \tag{5.17}$$

Après assemblage les équations (5.16) écrites pour tous les points de collocation (les nœuds du maillage par exemple) conduisent au système matriciel :

$$\mathbf{A}\mathbf{u} + \mathbf{B}\mathbf{q} = \mathbf{0} \tag{5.18}$$

Les conditions aux limites sont prises en compte en remplaçant les coordonnées de $\mathbf{u}$ et $\mathbf{q}$ correspondant par leurs valeurs dans le système (5.18).

3. dans cette expression, $u(y)$ et $(y) = \frac{\partial u(y)}{\partial n}$ sont les quantités interpolées

Calcul des intégrales

Les intégrales régulières sont calculées par intégration numérique de Gauss dans une ou dans deux directions.

Les intégrales singulières nécessitent un traitement particulier : intégration en valeur principale de Cauchy dont l'évaluation numérique correcte est maîtrisée de manière générale et satisfaisante depuis peu (1992 : Guiggiani et all).

5.4 Formulation indirecte : méthode des singularités

On considère le cas du problème du comportement d'un fluide parfait, incompressible en l'absence de distributions volumiques, et on repart de (4.5) dans laquelle, on identifie u au potentiel des vitesses et v à la fonction de Green G :

$$\int_\Omega (\mathrm{G}\Delta\Phi - \Phi\Delta\mathrm{G})\mathrm{d}\Omega = \int_{\partial\Omega} \left(\mathrm{G}\frac{\partial\Phi}{\partial n} - \Phi\frac{\partial \mathrm{G}}{\partial n}\right)\mathrm{d}\Gamma \tag{5.19}$$

Les fonctions qui satisfont l'équation locale c'est-à-dire l'équation de Laplace sont déterminées en utilisant la méthode des singularités.

5.4.1 Solutions de l'équation de Laplace

Pour construire la solution harmonique solution du problème de Neumann extérieur, on remplace la frontière $\partial\Omega$ par une distribution de singularités. Ces distributions peuvent être de différents types et en particulier :

- distribution de sources ;
- distribution de doublets normaux ;
- distribution mixte de Green.

Distribution de sources (potentiels de surface de simple couche)

(P) est une densité surfacique de source sur une surface S si le potentiel créé en M extérieur à S vaut :

$$\Phi(\mathrm{M}) = -\frac{1}{4\pi}\int_\mathrm{S} \frac{\sigma(\mathrm{P})}{r}\mathrm{dS_P} \tag{5.20}$$

et satisfait les conditions suivantes :

$$\begin{aligned} &\Phi(\mathrm{M}^+) - \Phi(\mathrm{M}^-) = 0; && \left.\frac{\partial\Phi}{\partial n}\right|_{\mathrm{M}^+} - \left.\frac{\partial\Phi}{\partial n}\right|_{\mathrm{M}^-} = \frac{\sigma(\mathrm{M})}{2}; \\ &\left.\frac{\partial\Phi}{\partial n}\right|_{\mathrm{M}^+} - \left.\frac{\partial\Phi}{\partial n}\right|_{\mathrm{M}^-} = \frac{\sigma(\mathrm{M})}{2}; && \left.\frac{\partial\Phi}{\partial n}\right|_{\mathrm{M}^+} - \left.\frac{\partial\Phi}{\partial n}\right|_{\mathrm{M}^-} = 0 \end{aligned} \tag{5.21}$$

avec les trois points M^+, M, M^- infiniment voisins, M à la frontière S et M^+ et M^- situés de part et d'autre de celle-ci comme illustré sur la figure 5.1.

Distribution de doublets normaux (potentiels de double couche)

(P) est une distribution de doublets normaux sur une surface S si le potentiel créé en M extérieur à S vaut :

$$\Phi(\mathrm{M}) = -\frac{1}{4\pi}\int \mu(\mathrm{P})\frac{\partial}{\partial n_\mathrm{P}}\left(\frac{1}{r}\right)\mathrm{dS_P} \tag{5.22}$$

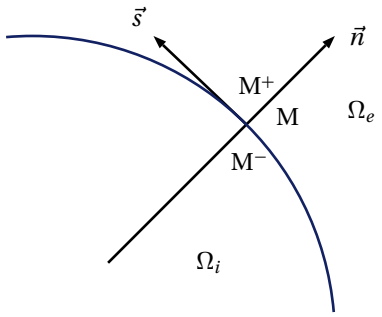


figure 5.1 - *Géométrie et notations à l'interface*

et satisfait les conditions suivantes :

$$\Phi(M^+)-\Phi(M)=-\frac{\mu(M)}{2};\qquad \Phi(M)-\Phi(M^-)=-\frac{\mu(M)}{2};\qquad \left(\frac{\partial\Phi}{\partial n}\right)_{M^+}-\left(\frac{\partial\Phi}{\partial n}\right)_{M^-}=0$$
$$\left(\frac{\partial\Phi}{\partial n}\right)_{M^+}-\left(\frac{\partial\Phi}{\partial n}\right)_{M}=-\frac{1}{2}\left(\frac{\partial\mu}{\partial s}\right)_{M};\qquad \left(\frac{\partial\Phi}{\partial n}\right)_{M}-\left(\frac{\partial\Phi}{\partial n}\right)_{M^-}=-\frac{1}{2}\left(\frac{\partial\mu}{\partial s}\right)_{M} \tag{5.23}$$

5.4.2 Applications : distribution mixte de Green

La méthode des singularités que nous utiliserons consiste à remplacer la frontière $\partial\Omega\equiv\Gamma_{SF}$ par une distribution mixte de Green : *distribution de sources (potentiels de surface de simple couche)* et *distribution de doublets normaux (potentiels de double couche).*

Distribution mixte de Green et domaine intérieur

Soit $\Phi(M)$ un champ scalaire harmonique défini dans le domaine intérieur Ω_i dont le domaine complémentaire est Ω_e et $\vec{n}_P$ la normale à la surface Γ_{SF} frontière des deux domaines, extérieure au domaine intérieur Ω_i. L'équation (5.19) écrite avec la fonction de Green $G(M,P)=\frac{1}{r}$ donne la troisième formule de Green qui s'écrit :

$$\int_{\Gamma_{SF}}\left(\Phi_i(P)\frac{\partial}{\partial n_P}\left(\frac{1}{r}\right)-\frac{1}{r}\frac{\partial\Phi_i(P)}{\partial n_P}\right)dS=\begin{cases}-4\pi\Phi_i(M) & \text{si } M\in\Omega_i\\ -2\pi\Phi_i(M) & \text{si } M\in\Gamma_{SF}\\ 0 & \text{si } M\in\Omega_e\end{cases} \tag{5.24}$$

Le potentiel en $M_i(M)$ apparaît comme la superposition des potentiels créés par :

- une densité surfacique de sources : $\sigma(P)=-\left(\frac{\partial\Phi_i}{\partial n}\right)_P$
- une densité surfacique de doublets normaux : $\mu(P)=\Phi_i(P)$

Distribution mixte de Green et domaine extérieur

Ces résultats sont encore applicables à un champ scalaire harmonique $\Phi_e(M)$, défini dans un domaine extérieur Ω'_e limité par Γ_{SF} et la surface d'une sphère entourant entièrement le domaine Ω_i dont le rayon tend vers l'infini.

Les termes d'intégration sur cette frontière sont nuls, une fonction harmonique, régulière, tend vers zéro à l'infini comme $1/r$. En choisissant comme normale à Γ_{SF} la nor-

male extérieure au domaine intérieur Ω_i, on obtient :

$$-\int_{\Gamma_{SF}} \left(\Phi_e(P) \frac{\partial}{\partial n_P} \left(\frac{1}{r} \right) - \frac{1}{r} \frac{\partial \Phi_e(P)}{\partial n_P} \right) dS = \begin{cases} 0 & \text{si } M \in \Omega_i \\ -2\pi\Phi_e(M) & \text{si } M \in \Gamma_{SF} \\ -4\pi\Phi_e(M) & \text{si } M \in \Omega_e \end{cases} \tag{5.25}$$

Distribution mixte de Green sur Γ_{SF}

En ajoutant les représentations dans Ω_i et Ω_e, on obtient :

$$\int_{\Gamma_{SF}} \left[\left(\Phi_e(P) - \Phi_i(P) \right) \frac{\partial}{\partial n_P} \left(\frac{1}{r} \right) - \frac{1}{r} \left(\frac{\partial \Phi_e(P)}{\partial n_P} - \frac{\partial \Phi_i(P)}{\partial n_P} \right) \right] dS = \begin{cases} 4\pi\Phi_i(M) & \text{si } M \in \Omega_i \\ 2\pi(\Phi_e(M) + \Phi_i(M)) & \text{si } M \in \Gamma_{SF} \\ 4\pi\Phi_e(M) & \text{si } M \in \Omega_e \end{cases} \tag{5.26}$$

Donc les densités surfaciques de singularité sur Γ_{SF} :

$$\begin{aligned} \sigma(P) &= \frac{\partial \Phi_e}{\partial n} - \frac{\partial \Phi_i}{\partial n} \\ \mu(P) &= -\Phi_e + \Phi_i \end{aligned} \tag{5.27}$$

induisent les potentiels Φ_e et Φ_i dans les domaines extérieur et intérieur et l'équation intégrale (5.26) s'écrit :

$$\int_{\Gamma_{SF}} \left(\mu(P) \frac{\partial}{\partial n_P} \left(\frac{1}{r} \right) + \frac{1}{r} \sigma(P) \right) dS = \begin{cases} -4\pi\Phi_i(M) & \text{si } M \in \Omega_i \\ -2\pi(\Phi_i(M) + \Phi_e(M)) & \text{si } M \in \Gamma_{SF} \\ -4\pi\Phi_e(M) & \text{si } M \in \Omega_e \end{cases} \tag{5.28}$$

5.5 Discrétisation et résolution

On cherche la distribution mixte de Green solution du problème intérieur ou extérieur.

5.5.1 Problème intérieur

On obtient la distribution mixte de Green en choisissant comme potentiel intérieur le potentiel Φ_e identiquement nul dans Ω_e et dans ces conditions la distribution surfacique de sources est définie sur Γ_{SF} par :

$$\begin{aligned} \mu(P) &= \Phi_i(P) \\ \sigma(P) &= -\frac{\partial \Phi_i}{\partial n} \end{aligned} \tag{5.29}$$

En écrivant l'équation (5.24) pour un point M appartenant à la frontière Γ_{SF} il vient :

$$-\frac{\mu(P)}{2} - \frac{1}{4\pi} \int_{\Gamma_{SF}} \mu(P) \frac{\partial}{\partial n_P} \left(\frac{1}{r} \right) dS = \frac{1}{4\pi} \int_{\Gamma_{SF}} \sigma(P) \frac{dS}{r} \tag{5.30}$$

Pour résoudre cette équation intégrale, la méthode de discrétisation adoptée est une méthode de collocation par sous-domaines. La surface Γ_{SF} est discrétisée en N facettes planes quadrilatères ou triangulaires d'aire S_j supportant des densités surfaciques de singularités constantes. Donc sur chaque facette j, $\mu(P)$ et $\sigma(P)$ restent égales à leur valeur moyenne μ_j et σ_j.

Dans ces conditions l'équation intégrale sur la surface complète précédente est transformée en une somme d'intégrales calculées sur la surface de chaque facette i qui s'écrit :

$$-\frac{\mu_i}{2}-\sum_{j=1}^{N}\mu_j\left(\frac{1}{4\pi}\int_{S_j}\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\mathrm{d}S\right)=\sum_{j=1}^{N}\sigma_j\left(\frac{1}{4\pi}\int_{S_j}\frac{1}{r}\mathrm{d}S\right) \tag{5.31}$$

avec $r=|\mathrm{M}_i\mathrm{P}|$, et M_i centre de la facette i et le problème fluide discrétisé se ramène à la résolution d'un système d'équations linéaires :

$$\mathbf{D}_i\boldsymbol{\mu}=\mathbf{S}_i\boldsymbol{\sigma} \tag{5.32}$$

avec $\mathbf{D}_i$, matrice des coefficients d'influence de doublets normaux, telle que :

$$(d_{ij})_i=-\frac{\delta_i^j}{2}-\frac{1}{4\pi}\int_{S_j}\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\mathrm{d}S$$

et $\mathbf{S}_i$, matrice des coefficients d'influence de source, telle que :

$$(s_{ij})_i=\frac{1}{4\pi}\int_{S_j}\frac{\mathrm{d}S}{r}$$

5.5.2 Problème extérieur

On obtient la distribution mixte de Green en choisissant maintenant, comme potentiel intérieur le potentiel Φ_i identiquement nul dans Ω_i et dans ces conditions la distribution surfacique de sources est définie sur Γ_{SF} par :

$$\begin{aligned}\mu(\mathrm{P})&=-\Phi_e(\mathrm{P})\\ \sigma(\mathrm{P})&=\frac{\partial\Phi_e}{\partial n}\end{aligned} \tag{5.33}$$

L'équation intégrale (5.28) écrite pour un point M appartenant à la frontière Γ_{SF} est :

$$\frac{\mu(\mathrm{P})}{2}-\frac{1}{4\pi}\int_{\Gamma_{SF}}\left(\mu(\mathrm{P})\frac{\partial}{\partial n_\mathrm{P}}\left(\frac{1}{r}\right)\right)\mathrm{d}S=\frac{1}{4\pi}\int_{\Gamma_{SF}}\left(\sigma(\mathrm{P})\frac{1}{r}\right)\mathrm{d}S \tag{5.34}$$

et après discrétisation en facettes planes, il vient :

$$\frac{\mu_i}{2}-\sum_{j=1}^{N}\mu_j\left(\frac{1}{4\pi}\int_{S_j}\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\mathrm{d}S\right)=\sum_{j=1}^{N}\sigma_j\left(\frac{1}{4\pi}\int_{S_j}\frac{1}{r}\mathrm{d}S\right) \tag{5.35}$$

et le problème fluide discrétisé se ramène à la résolution d'un système d'équations linéaires :

$$\mathbf{D}_e\boldsymbol{\mu}=\mathbf{S}_e\boldsymbol{\sigma} \tag{5.36}$$

avec :

$$(d_{ij})_e=\frac{\delta_i^j}{2}-\frac{1}{4\pi}\int_{S_j}\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\mathrm{d}S\qquad(s_{ij})_e=\frac{1}{4\pi}\int_{S_j}\frac{\mathrm{d}S}{r}$$

Sur Γ_{SF}, $\sigma(\mathrm{P})$ s'identifie à la vitesse normale de la paroi, soit :

$$\frac{\partial\Phi}{\partial n}=\vec{\mathrm{V}}_s\cdot\vec{n} \tag{5.37}$$

et $\mu(P)$ au potentiel des vitesses. La connaissance de l'une ou l'autre de ces grandeurs ou une combinaison linéaire des deux, suffit pour résoudre complètement le problème.

Les résultats précédents sont généralisables à la recherche de fonctions scalaires satisfaisant l'équation de Helmoltz d'une part lorsque l'hypothèse de fluide compressible est retenue ou l'équation de Laplace d'autre part lorsque le fluide incompressible est limité par une surface libre déformable (ondes de gravité). Dans l'équation intégrale précédente, la fonction $1/r$ est remplacée par la fonction de Green $G(M,P)$ du problème correspondant. La forme intégrale $W(\Phi)$ devient une intégrale de Fredholm de deuxième espèce dont l'expression générale est :

$$\frac{\mu(P)}{2} - \frac{1}{4\pi}\int_{\Gamma_{SF}} \mu(P)\frac{\partial G(M,P)}{\partial n_P}dS = \frac{1}{4\pi}\int_{\Gamma_{SF}} \sigma(P)G(M,P)dS \tag{5.38}$$

6

Problèmes d'élastostatique

6.1 Théorème de réciprocité de Maxwell-Betti

L'identité de réciprocité pour l'élasticité est une conséquence du principe des travaux virtuels.

Quel que soit le champ de contrainte en équilibre statique avec f, quel que soit le déplacement virtuel $u'(u)$ sur :

$$\int_\Omega \sigma : \varepsilon(\vec{u}')\mathrm{dV} - \int_{\partial\Omega} (\sigma\cdot\vec{n})\cdot\vec{u}'\mathrm{dS} - \int_\Omega \rho\vec{f}\cdot\vec{u}'\mathrm{dV} = 0 \tag{6.1}$$

En écrivant le principe des travaux virtuels pour deux états élastostatique (u^1,σ^1,f^1) et (u^2,σ^2,f^2), et compte tenu de la symétrie de la loi de comportement, il en résulte que tout couple (u^1,σ^1,f^1), (u^2,σ^2,f^2) vérifie l'identité de Maxwell-Betti :

$$\int_{\partial\Omega} \left[(\sigma^1\cdot\vec{n})\cdot\vec{u}^2 - (\sigma^2\cdot\vec{n})\cdot\vec{u}^1\right]\mathrm{dS} = \int_\Omega \left[\rho\vec{f}^2\cdot\vec{u}^1 - \rho\vec{f}^1\cdot\vec{u}^2\right]\mathrm{dV}(3-2) \tag{6.2}$$

qui peut également s'écrire compte tenu de l'équation de Lamé-Navier :

$$\int_{\partial\Omega} \left[\mathscr{B}(u^1)\cdot\vec{u}^2 - \mathscr{B}(u^2)\cdot\vec{u}^1\right]\mathrm{dS} = \int_\Omega \left[(\Delta^* u^1)\cdot\vec{u}^2 - (\Delta^* u^2)\cdot\vec{u}^1\right]\mathrm{dV} \tag{6.3}$$

On retrouve la structure générale de (4.6).

6.2 Solutions élémentaires de l'élasticité linéaire

On appelle solution élémentaire un état élastostatique $(\mathrm{U}^k,\Sigma^k,\mathrm{F}^k)$ associé à une force ponctuelle unitaire appliquée en un point x fixé et de direction $\vec{e}_k$:

$$\rho\mathrm{F}^k(y) = \delta(y-x)e_k \qquad y\in\Omega \tag{6.4}$$

– pour *l'espace infini*, la solution élémentaire est donnée par la solution de Kelvin :

$$\begin{aligned}
\mathrm{U}_i^k(x,y) &= \frac{1}{16\pi\mu(1-\nu)r}\left[r_{,i}\,r_{,k} + (3-4\nu)\delta_{ik}\right] \\
\Sigma_{ij}^k(x,y) &= \frac{1}{8\pi(1-\nu)r^2}\left[3r_{,i}\,r_{,k}\,r_{,j} + (1-2\nu)(\delta_{ik}r_{,j} + \delta_{jk}r_{,i} - \delta_{ij}r_{,k})\right]
\end{aligned} \tag{6.5}$$

où $r = |r| = \left|y-x\right|$

– pour *le demi espace avec surface libre*, la solution élémentaire est donnée par la solution de Mindlin correspondant à une force ponctuelle appliquée en un point intérieur. Elle contient comme cas particulier la solution de Boussinesq correspondant à une force ponctuelle appliquée normalement à la surface libre.

Le traitement se fait suivant la démarche générale donnée au chapitre précédent dans le cas de l'équation de Poisson, les difficultés étant toutefois augmentées.

Troisième partie

DIFFÉRENCES FINIES

Yves Lecointe

7

Principes généraux

Le but de ce chapitre est de présenter les bases nécessaires à l'utilisation des méthodes de différences finies ; c'est pourquoi nous nous limitons aux problèmes unidirectionnels, donc à des problèmes à une seule variable. Les problèmes différentiels traités sont des problèmes du second ordre appelés souvent laplacien unidirectionnel. Physiquement, un laplacien unidirectionnel correspond à des phénomènes stationnaires (indépendants du temps), de type diffusion. On retrouve ainsi l'équation de la chaleur stationnaire.

7.1 Introduction

7.1.1 Principe

On considère un problème différentiel d'ordre 2 d'une fonction à une variable x. Soit une fonction u supposée continue et dérivable sur un intervalle $[x_a, x_b]$ et un opérateur différentiel $\mathscr{L}$ défini par :

$$\mathscr{L}(u) = a\,u'' + b\,u' + c\,u \tag{7.1}$$

autrement dit :

$$\mathscr{L} = a\frac{\mathrm{d}^2}{\mathrm{d}x^2} + b\frac{\mathrm{d}}{\mathrm{d}x} + c \tag{7.2}$$

où les coefficients a, b et c sont des fonctions données de la variable x. Cet opérateur est appelé aussi opérateur elliptique en dimension 1. D'un point de vue physique, ce type d'équation différentielle correspond à la modélisation d'un problème de diffusion $a\,u''$ où a est le coefficient de diffusion de la quantité u, combinée avec de l'advection $b\,u'$, où b est une vitesse de transport de la quantité u. Lorsque b est nul, nous avons affaire à un problème de diffusion pure. Le terme $c\,u$ est appelé terme *source* car il correspond à la production (ou à la destruction) de la quantité u.

On pose le problème suivant (problème de Dirichlet-Dirichlet) :

$$\begin{aligned} &\mathscr{L}(u) = f \quad \text{sur} \quad]x_a, x_b[\\ &u(x_a) = u_a \\ &u(x_b) = u_b \end{aligned} \tag{7.3}$$

ou le problème suivant (Problème de Dirichlet-Neumann) :

$$\begin{aligned} &\mathscr{L}(u) = f \quad \text{sur} \quad]x_a, x_b[\\ &u(x_a) = u_a \quad \text{ou} \quad u'(x_a) = u'_a \\ &u'(x_b) = u'_b \qquad\quad u(x_b) = u_b \end{aligned} \tag{7.4}$$

On cherche à résoudre ces problèmes de façon approchée, en remplaçant dans ces équations, les dérivées par des formules approchées qui feront intervenir des différences entre des valeurs approchées discrètes de la fonction $u(x)$. On notera $\mathrm{U}_j \simeq u(x_j)$, la valeur discrète approchée de la fonction u au point d'abscisse x_j, $u(x_j)$ étant la valeur exacte de la solution du problème.

7.1.2 Exemples

Barre verticale

Soit une barre verticale de longueur ℓ encastrée et soumise à son poids propre. L'équation différentielle régissant les variations du déplacement $u(x)$ s'écrit :

$$\mathrm{ES}\frac{\mathrm{d}^2u}{\mathrm{d}x^2} + p = 0 \tag{7.5}$$

où E est le module d'Young du matériau, S, la section de la barre et p, la charge linéique. Si on ne considère que la charge due au poids propre, $p = \ell g \mathrm{S}$. À l'encastrement, la condition limite est $u(0) = 0$, et à l'autre extrémité $x = \ell$, on peut écrire une condition sur l'effort normal :

$$\mathrm{N}(\ell) = \mathrm{ES}\left.\frac{\mathrm{d}u}{\mathrm{d}x}\right|_{\ell} = \mathrm{F} \tag{7.6}$$

où la force de traction F peut être nulle. Dans le cas où la barre est soumise à son poids propre, l'équation :

$$\frac{\mathrm{d}^2u}{\mathrm{d}x^2} + \frac{\rho g}{\mathrm{E}} = 0 \tag{7.7}$$

a pour solution :

$$u(x) = -\frac{\rho g}{2\mathrm{E}}x(x - 2\ell) \tag{7.8}$$

Si en plus, cette barre est en traction sous effort extérieur, la solution est :

$$u(x) = -\frac{\rho g}{2\mathrm{E}}x\left(x - 2\ell + \frac{\mathrm{F}}{\rho g}\right) \tag{7.9}$$

Conduction dans une barre

Soit l'équation différentielle régissant les variations de la température d'une barre de longueur ℓ :

$$a\frac{\mathrm{d}^2\mathrm{T}}{\mathrm{d}x^2} + q = 0 \tag{7.10}$$

où a est le coefficient de diffusion de la chaleur du matériau. À l'extrémité $x = 0$, la condition limite est $\mathrm{T}(0) = \mathrm{T}_0$, et à l'autre extrémité ($x = \ell$), on peut écrire soit une condition de Dirichlet $u(\ell) = \mathrm{T}_\ell$, soit une condition de flux (de type Neumann) $\mathrm{F}(\ell) = k\frac{\mathrm{dT}}{\mathrm{d}x}$, qui est soit nul, soit égal à une valeur Φ_1.

Advection diffusion

On considère un tube de rayon R et de longueur L, dans lequel circule un fluide. La vitesse et la température du fluide sont supposées uniformes dans la section, ce qui permet de considérer que le phénomène est régi par une équation monodimensionnelle. Le

transfert de chaleur convectif entre le fluide et la paroi du tube se fait avec un coefficient de transfert égal à h_c :

$$h_c p \mathrm{R}^2 r \mathrm{C}_p \mathrm{V} \frac{\mathrm{dT}}{\mathrm{d}x} + 2p\mathrm{R}h_c(\mathrm{T}-\mathrm{T}_0) = p\mathrm{R}^2 k \frac{\mathrm{d}^2\mathrm{T}}{\mathrm{d}x^2} \tag{7.11}$$

On suppose que la température à l'entrée du tube $x = x_0$ est égale à T_0, alors qu'elle est égale à T_L, à la sortie $x = x_\mathrm{L}$. Afin de trouver la solution analytique de ce problème, on travaillera avec des variables réduites, $\mathrm{T}' = (\mathrm{T}-\mathrm{T}_0)/(\mathrm{T}_\mathrm{L}-\mathrm{T}_0)$ et $x' = x/\mathrm{R}$. L'équation transformée est :

$$\frac{\mathrm{d}^2\mathrm{T}'}{\mathrm{d}x'^2} - \frac{\mathrm{R}r\mathrm{C}_p\mathrm{V}}{k}\frac{\mathrm{dT}'}{\mathrm{d}x'} - \frac{2\mathrm{R}h_c}{k}\mathrm{T}' = 0 \tag{7.12}$$

Dans cette équation, nous avons deux coefficients adimensionnels, $\mathrm{R}r\mathrm{C}_p\mathrm{V}/k$ et $\mathrm{R}h_c/k$. On pose $\mathrm{L} = 5\mathrm{R}$ et $\mathrm{R}r\mathrm{C}_p\mathrm{V}/k = \mathrm{R}h_c/k = 1$; le problème précédent s'écrit :

$$\begin{aligned} &\frac{\mathrm{d}^2\mathrm{T}'}{\mathrm{d}x'^2} - \frac{\mathrm{dT}'}{\mathrm{d}x'} - 2\mathrm{T}' = 0 \quad \text{sur} \quad]0,10[\\ &\mathrm{T}'(0) = 0 \\ &\mathrm{T}'(10) = 1 \end{aligned} \tag{7.13}$$

La solution analytique est de la forme $\mathrm{T}'(x') = \alpha e^{-x'} + \beta e^{2x'}$, avec α et β déterminés à partir des conditions limites, soit :

$$\mathrm{T}'(x') = \frac{e^{-x'} - e^{2x'}}{e^{-5} - e^{10}} \tag{7.14}$$

7.1.3 Approximation de Lagrange en trois points

On considère les trois points d'abscisses x_{j-1}, x_j et x_{j+1} et on cherche à évaluer la valeur de la dérivée première et de la dérivée seconde de la fonction $u(x)$ en fonction des trois valeurs discrètes U_{j-1}, U_j et U_{j+1}.

Dérivée première

Si f est une fonction continue, la définition classique de la dérivée à droite au point x_j est :

$$u'(x_j) = \lim_{x_{j+1}\to x_j} \frac{u(x_{j+1}) - u(x_j)}{x_{j+1} - x_j} \tag{7.15}$$

on peut aussi définir la dérivée à gauche par :

$$u'(x_j) = \lim_{x_{j-1}\to x_j} \frac{u(x_j) - u(x_{j-1})}{x_j - x_{j-1}} \tag{7.16}$$

Si maintenant on considère les valeurs discrètes, nous aurons les formules approchées suivantes :

$$\begin{aligned} u'(x_j) &\simeq \frac{\mathrm{U}_{j+1} - \mathrm{U}_j}{x_{j+1} - x_j} = \frac{\mathrm{U}_{j+1} - \mathrm{U}_j}{h_{j+1}} \qquad \text{dérivée avant (forward)} \\ u'(x_j) &\simeq \frac{\mathrm{U}_j - \mathrm{U}_{j-1}}{x_j - x_{j-1}} = \frac{\mathrm{U}_j - \mathrm{U}_{j-1}}{h_j} \qquad \text{dérivée arrière (backward)} \end{aligned} \tag{7.17}$$

où $h_{j+1} = x_{j+1} - x_j$ et $h_j = x_j - x_{j-1}$. En combinant ces deux dernières formules, nous obtenons :

$$u'(x_j) \simeq \frac{\theta}{h_{j+1}}(U_{j+1} - U_j) + \frac{1-\theta}{h_j}(U_j - U_{j-1}) \tag{7.18}$$

soit encore, en réduisant au même dénominateur :

$$u'(x_j) \simeq \frac{\theta h_j U_{j+1} + [(1-\theta)h_{j+1} - \theta h_j]U_j - (1-\theta)h_{j+1}U_{j-1}}{h_{j+1}h_j} \tag{7.19}$$

De cette expression découlent plusieurs cas :

- si $\theta = 1/2$:

$$u'(x_j) \simeq \frac{h_j U_{j+1} + [h_{j+1} - h_j]U_j - h_{j+1}U_{j-1}}{2h_{j+1}h_j} \tag{7.20}$$

- si $\theta = \frac{h_{j+1}}{h_{j+1}+h_j}$, le terme central en U_j disparaît :

$$u'(x_j) \simeq \frac{U_{j+1} - U_{j-1}}{h_{j+1} + h_j} \tag{7.21}$$

- si le pas est constant, alors $h = h_j = h_{j+1}$:

$$u'(x_j) \simeq \frac{\theta U_{j+1} + (1-2\theta)U_j - (1-\theta)U_{j-1}}{h} \tag{7.22}$$

  - $\theta = 1$: dérivée décentrée à droite ;
  - $\theta = 0$: dérivée décentrée à gauche ;
  - $\theta = 1/2$, la formule (7.22) s'écrit :

$$u'(x_j) \simeq \frac{U_{j+1} - U_{j-1}}{2h} \tag{7.23}$$

    dérivée première centrée et le point j n'intervient plus ; cette formule est appelée formule centrée par opposition aux formules précédentes qui sont des formules décentrées à gauche ou à droite.

La figure 7.1 donne une interprétation géométrique de ces trois formules de dérivation.

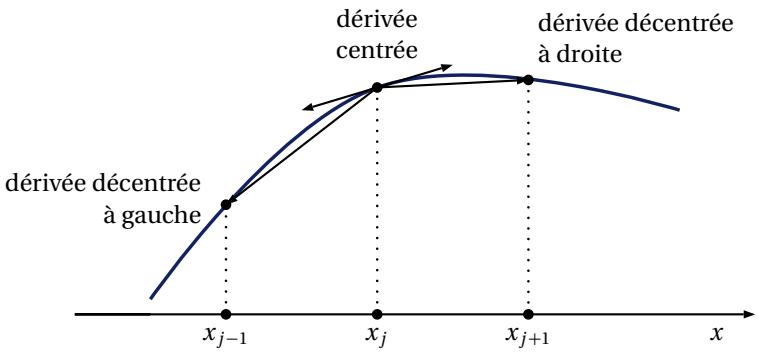


figure 7.1 - *Différents types de dérivée première*

Dérivée seconde

Soient $x_{j+1/2}$ et $x_{j-1/2}$, les milieux respectifs des segments $[x_{j+1};x_j]$ et $[x_j;x_{j-1}]$, pour calculer la dérivée seconde de $u(x)$ au point j, nous écrivons :

$$u''(x_j) \simeq \frac{u'(x_{j+1/2}) - u'(x_{j-1/2})}{x_{j+1/2} - x_{j-1/2}} \tag{7.24}$$

soit en remplaçant $u'(x_{j+1/2})$ et $u'(x_{j-1/2})$ par les formules approchées suivantes :

$$u'(x_{j+1/2}) = \frac{\mathrm{U}_{j+1} - \mathrm{U}_j}{h_{j+1}} \quad \text{et} \quad u'(x_{j-1/2}) = \frac{\mathrm{U}_j - \mathrm{U}_{j-1}}{h_j} \tag{7.25}$$

nous obtenons :

$$u''(x_j) \simeq \frac{2(h_j \mathrm{U}_{j+1} - (h_{j+1}+h_j)\mathrm{U}_j + h_{j+1}\mathrm{U}_{j-1})}{h_{j+1}h_j(h_{j+1}+h_j)} \tag{7.26}$$

et, dans le cas d'un pas constant :

$$u''(x_j) \simeq \frac{\mathrm{U}_{j+1} - 2\mathrm{U}_j + \mathrm{U}_{j-1}}{h^2} \tag{7.27}$$

qui est la formule de la dérivée seconde centrée.

> Les deux formules (7.25) sont des formules d'approximation de la dérivée première, centrées au point $x_{j+1/2}$ ou au point $x_{j-1/2}$.

Erreur locale d'approximation

Nous voyons que dans le calcul de la dérivée première à trois points, si le pas est constant, il y a trois formules possibles ; évaluons la précision de chacune de ces formules. Pour cela, nous utilisons les développements en série de Taylor :

$$\begin{aligned} \mathrm{U}_{j-1} &\simeq u(x_{j-1}) \simeq u(x_j) - hu'(x_j) + \frac{h^2}{2!}u''(x_j) + \ldots + (-1)^n \frac{h^n}{n!}u^{(n)}(x_j) \\ \mathrm{U}_{j+1} &\simeq u(x_{j+1}) \simeq u(x_j) + hu'(x_j) + \frac{h^2}{2!}u''(x_j) + \ldots + \frac{h^n}{n!}u^{(n)}(x_j) \end{aligned} \tag{7.28}$$

Pour évaluer l'erreur d'une formule d'approximation, il suffit de remplacer les valeurs discrètes par les valeurs continues et d'effectuer les développements en série de Taylor. Si la formule de discrétisation en trois points a pour forme générale :

$$\mathrm{T}_j = r^+\mathrm{U}_{j+1} + r^c\mathrm{U}_j + r^-\mathrm{U}_{j-1} \tag{7.29}$$

le développement en série de Taylor peut se calculer et nous aurons :

$$\mathrm{T}_j = \alpha_0 u(x_j) + \sum_{i=1}^{\infty} \alpha_i h^i u^{(k)}(x_j) \tag{7.30}$$

la somme est appelée erreur de troncature et l'ordre de cette erreur est donné par le premier coefficient α_i non nul de la série.

> L'ordre de l'erreur de troncature correspond à l'exposant i du pas h. l'ordre de la dérivée k n'est donc pas celui de l'erreur.

Exemples

Pas constant

Si on reprend les trois formules suivantes, on peut évaluer l'erreur de troncature de chacune de ces trois formules en utilisant la méthode précédente :

$$\begin{aligned} \mathrm{T} &= \frac{\theta \mathrm{U}_{j+1} + (1-2\theta)\mathrm{U}_j - (1-\theta)\mathrm{U}_{j-1}}{h} \\ &\simeq \frac{\theta u(x_{j+1}) + (1-2\theta)u(x_j) - (1-\theta)u(x_{j-1})}{h} \end{aligned} \tag{7.31}$$

or :

$$\begin{aligned} \frac{\theta u(x_{j+1}) + (1-2\theta)u(x_j) - (1-\theta)u(x_{j-1})}{h} &= \frac{\theta + (1-2\theta) - (1-\theta)}{h} u(x_j) \\ &+ h\big(\theta + (1-\theta)\big)u'(x_j) + \frac{h^2}{2}\big(\theta - (1-\theta)\big)u''(x_j) + \frac{h^3}{3!}\big(\theta + (1-\theta)\big)u'''(x_j) + \mathcal{O}(h^4) \end{aligned}$$

soit encore :

$$\mathrm{T} \simeq u'(x_j) + \frac{h}{2}(2\theta - 1)u''(x_j) + \frac{h^2}{6}u'''(x_j) + \mathcal{O}(h^3) \tag{7.32}$$

Si $\theta = 1/2$, le premier terme de l'erreur de troncature sera du second ordre et donc une formule centrée sera plus précise qu'une formule décentrée, du premier ordre. Ainsi on peut écrire :

$$\begin{aligned} \frac{\mathrm{U}_{j+1} - \mathrm{U}_j}{h} &\simeq u'(x_j) + \frac{h}{2}u''(x_j) \\ \frac{\mathrm{U}_j - \mathrm{U}_{j-1}}{h} &\simeq u'(x_j) - \frac{h}{2}u''(x_j) \\ \frac{\mathrm{U}_{j+1} - \mathrm{U}_{j-1}}{2h} &\simeq u'(x_j) + \frac{h^2}{6}u'''(x_j) \end{aligned} \tag{7.33}$$

Pour réduire l'erreur d'approximation, la seule solution consiste à utiliser plus de points d'appui pour approximer les dérivées, par exemple 4 ou 5 points.

Pas variable

La formule générale d'une dérivée première en trois points est :

$$u'(x_j) \simeq \frac{\theta h_j \mathrm{U}_{j+1} + \big((1-\theta)h_{j+1} - \theta h_j\big)\mathrm{U}_j - (1-\theta)h_{j+1}\mathrm{U}_{j-1}}{h_{j+1}h_j} \tag{7.34}$$

en effectuant un développement en séries de Taylor autour du point x_j, nous aurons :

$$\begin{aligned} &\frac{1}{h_{j+1}h_j}\Big(\theta h_j \mathrm{U}_{j+1} + \big((1-\theta)h_{j+1} - \theta h_j\big)\mathrm{U}_j - (1-\theta)h_{j+1}\mathrm{U}_{j-1}\Big) = \\ &\qquad u'(x_j) + \frac{u''(x_j)}{2h_{j+1}h_j}\Big(\theta h_j h_{j+1}^2 - (1-\theta)h_{j+1}h_j^2\Big) \\ &\qquad + \frac{u'''(x_j)}{6h_{j+1}h_j}\Big(\theta h_j h_{j+1}^3 + (1-\theta)h_{j+1}h_j^3\Big) \end{aligned} \tag{7.35}$$

soit encore :

$$\frac{1}{h_{j+1}h_j}\Big(\theta h_j \mathrm{U}_{j+1}+\big((1-\theta)h_{j+1}-\theta h_j\big)\mathrm{U}_j-(1-\theta)h_{j+1}\mathrm{U}_{j-1}\Big)= \\ u'(x_j)+\frac{u''(x_j)}{2}\big(\theta h_{j+1}-(1-\theta)h_j\big)+\frac{u'''(x_j)}{6}\left(\theta h_{j+1}^2+(1-\theta)h_j^2\right) \tag{7.36}$$

d'où la discussion suivante :

- $\theta=0$: l'erreur de troncature est $-\frac{u''(x_j)}{2}h_j+\frac{u'''(x_j)}{6}h_j^2$, en $\mathcal{O}(h^j)$
- $\theta=1$: l'erreur de troncature est $\frac{u''(x_j)}{2}h_{j+1}-\frac{u'''(x_j)}{6}h_j^2$, en $\mathcal{O}(h_{j+1})$
- $\theta=1/2$: l'erreur de troncature est $\frac{u''(x_j)}{4}(h_{j+1}-h_j)+\frac{u'''(x_j)}{12}(h_j^2+h_{j+1}^2)$, en $\mathcal{O}(h_{j+1}-h_j)$
- $\theta=\frac{h_j}{h_{j+1}+h_j}$: l'erreur de troncature est $\frac{u''(x_j)}{2}(h_{j+1}h_j)$

Globalement, pour des valeurs $\theta\neq\frac{h_j}{h_{j+1}+h_j}$, ces formules sont en $\mathcal{O}(h)$, avec $h=\max[h_{j+1},h_j]$. Il est cependant possible de trouver autrement la formule centrée en pas variable ayant la précision la plus élevée, en partant de la forme :

$$u'(x_j)\simeq\gamma\mathrm{U}_{j+1}+\beta\mathrm{U}_j+\alpha\mathrm{U}_{j-1} \tag{7.37}$$

Les coefficients α, β et γ sont déterminés par les trois relations :

$$\left.\begin{aligned}\alpha+\beta+\gamma&=0\\ -\alpha h_j+\gamma h_{j+1}&=1\\ \alpha h_j^2+\gamma h_{j+1}^2&=0\end{aligned}\right\}\Leftrightarrow\begin{cases}\alpha=-\dfrac{h_{j+1}}{h_j(h_{j+1}+h_j)}\\ \beta=\dfrac{h_{j+1}-h_j}{h_jh_{j+1}}\\ \gamma=\dfrac{h_j}{h_{j+1}(h_{j+1}+h_j)}\end{cases} \tag{7.38}$$

et la formule résultante s'écrit :

$$u'(x_j)\simeq\frac{-h_{j+1}^2\mathrm{U}_{j-1}+h_{j+1}^2-h^2j\mathrm{U}_j+h_j^2\mathrm{U}_{j+1}}{h_{j+1}h_j(h_{j+1}+h_j)} \tag{7.39}$$

et l'erreur de troncature est égale à $u'''(x_j)\frac{-\alpha h_j^3+\gamma h_{j+1}^3}{3!}$, soit $u'''(x_j)\frac{h_{j+1}h_j}{6}$. Cette formule est d'ordre $\mathcal{O}(h_{j+1}h_j)$ et la précision est celle trouvée précédemment pour θ tel que le coefficient de la dérivée seconde était nul.

7.1.4 Approximation de Lagrange en n points

On considère les points d'abscisses x_j pour $j=1,n$. La formule la plus générale liant ces points peut s'écrire sous la forme :

$$\mathrm{T}_j=\sum_{i=-i_1}^{i_2}\alpha_i\mathrm{U}_{j+i}=\sum_{i=1}^{i_1}\alpha_i^-\mathrm{U}_{j-i}+\sum_{i=0}^{i_2}\alpha_i\mathrm{U}_{j+i}=\sum_{i=-i_1}^{-1}\alpha_i^-\mathrm{U}_{j+i}+\sum_{i=0}^{i_2}\alpha_i\mathrm{U}_{j+i} \tag{7.40}$$

cette formule contient (i_1+i_2+1) coefficients avec $(i_1+i_2+1)\leqslant n$. Pour déterminer les coefficients α_i, effectuons le développement en séries de Taylor de T_j :

$$\mathrm{T}_j=\sum_{i=0}^{n}\beta_i u^{(i)}(x_j) \tag{7.41}$$

Par identification, il vient :

$$\begin{aligned}
u'(x_j)\beta_1 &= \sum_{i=-i_1}^{i_2} \alpha_i u'(x_j) = \sum_{i=-i_1}^{-1} \alpha_i^- i h + \sum_{i=1}^{i_2} \alpha_i i h = h\left(\sum_{i=-i_1}^{-1} \alpha_i^- i + \sum_{i=1}^{i_2} \alpha_i i\right) \\
u''(x_j)\beta_k &= \frac{1}{2}\left(\sum_{i=-i_1}^{-1} \alpha_i^- (ih)^2 + \sum_{i=1}^{i_2} (\alpha_i h)^2\right) = \frac{h^2}{k!}\left(\sum_{i=-i_1}^{-1} \alpha_i^- i^2 + \sum_{i=1}^{i_2} \alpha_i i^2\right) \\
u^{(k)}(x_j)\beta_k &= \frac{1}{k!}\left(\sum_{i=-i_1}^{-1} \alpha_i^- (ih)^k + \sum_{i=1}^{i_2} (\alpha_i h)^k\right) = \frac{h^k}{k!}\left(\sum_{i=-i_1}^{-1} \alpha_i^- i^k + \sum_{i=1}^{i_2} \alpha_i i^k\right)
\end{aligned} \tag{7.42}$$

Pour déterminer la dérivée d'ordre m, il suffit de calculer les coefficients α_i correspondant à $\beta_i = \delta_{im}$ pour $i = -i_1,\ i_2$. Il faut bien résoudre un système de (i_1+i_2+1) équations à (i_1+i_2+1) inconnues.

Dérivées premières en trois points décentrées

Si on cherche la formule donnant, en pas constant, la dérivée première au point j en fonction des points $j-1$ et $j-2$, il nous faut trois équations ; en annulant les coefficients de u et de u'' et en prenant celui de u' égal à 1, nous avons 3 équations à 3 inconnues et nous avons la formule suivante :

$$u'(x_j) \simeq \frac{3\mathrm{U}_j - 4\mathrm{U}_{j-1} + \mathrm{U}_{j-2}}{2h} \tag{7.43}$$

et la formule similaire faisant intervenir les points $j+1$ et $j+2$:

$$u'(x_j) \simeq \frac{-3\mathrm{U}_j + 4\mathrm{U}_{j+1} - \mathrm{U}_{j+2}}{2h} \tag{7.44}$$

7.1.5 Approximations polynomiales

Dans les paragraphes précédents, nous avons considéré les développements en séries de Taylor d'une fonction et les formules permettant d'approcher les dérivées premières et secondes ont été construites à partir de ces développements. Une autre approche consiste à utiliser des approximations polynomiales d'une fonction (voir chapitre correspondant). Il est évident que ces deux approches sont similaires et conduisent aux mêmes formules. Ainsi, si on considère une fonction $u(x)$ définie en trois points x_{j-1}, x_j et x_{j+1}, cette fonction peut être approximée par le polynôme du second degré (approximation de type parabolique) :

$$\mathrm{U}(x) = ax^2 + bx + c \tag{7.45}$$

et les coefficients a, b et c seront déterminés de façon à avoir :

$$\begin{aligned}
\mathrm{U}_{j-1} &= a\big((j-1)h\big)^2 + b(j-1)h + c \\
\mathrm{U}_j &= a\big(jh\big)^2 + bjh + c \\
\mathrm{U}_{j+1} &= a\big((j+1)h\big)^2 + b(j+1)h + c
\end{aligned} \tag{7.46}$$

La dérivée première sera donc $U'(x) = 2ax + b$ et la dérivée seconde $U''(x) = 2a$. Les coefficients a, b et c sont :

$$\begin{aligned} a &= \frac{U_{j+1} - 2U_j + U_{j-1}}{2h^2} \\ b &= \frac{-(2j-1)U_{j+1} + 4jU_j - (2j+1)U_{j-1}}{2h} \\ c &= \frac{j(j-1)U_{j+1} - 2(j^2-1)U_j + j(j+1)U_{j-1}}{2} \end{aligned} \tag{7.47}$$

On vérifie que :

$$\begin{aligned} U'(x_{j-1}) &= 2a(j-1)h + b = \frac{-3U_{j-1} + 4U_j - U_{j+1}}{2h} \\ U'(x_j) &= 2aj\,h + b = \frac{U_{j+1} - U_{j-1}}{2h} \\ U'(x_{j+1}) &= 2a(j+1)h + b = \frac{U_{j-1} - 4U_j + 3U_{j+1}}{2h} \end{aligned} \tag{7.48}$$

et de la même façon, pour la dérivée seconde :

$$U''(x_{j-1}) = U''(x_j) = U''(x_{j+1}) = \frac{U_{j+1} - 2U_j + U_{j-1}}{2h^2} \tag{7.49}$$

Il est clair que si on utilise seulement deux points (approximation linéaire) le polynôme sera de la forme $U(x) = a'x + b'$ et la dérivée première sera $U'(x) = a'$ et la dérivée seconde sera nulle.

7.2 Formulaire

7.2.1 Dérivées premières décentrées

Formules à deux points

$$\begin{aligned} u'(x_j) &\simeq \frac{U_{j+1} - U_j}{h} + \mathcal{O}(h) \\ u'(x_j) &\simeq \frac{U_j - U_{j-1}}{h} + \mathcal{O}(h) \end{aligned} \tag{7.50}$$

Formules à trois points

$$\begin{aligned} u'(x_j) &\simeq \frac{-3U_j + 4U_{j+1} - U_{j+2}}{2h} + \mathcal{O}(h^2) \\ u'(x_j) &\simeq \frac{3U_j - 4U_{j-1} + U_{j-2}}{2h} + \mathcal{O}(h^2) \end{aligned} \tag{7.51}$$

Formules à quatre points

$$\begin{aligned} u'(x_j) &\simeq \frac{U_{j-2} - 6U_{j-1} + 3U_j + 2U_{j+1}}{6h} + \mathcal{O}(h^3) \\ u'(x_j) &\simeq \frac{-2U_{j-1} - 3U_j + 6U_{j+1} - U_{j+2}}{6h} + \mathcal{O}(h^3) \\ u'(x_j) &\simeq \frac{-11U_j + 18U_{j+1} - 9U_j + 2 + 2U_{j+3}}{6h} + \mathcal{O}(h^3) \end{aligned} \tag{7.52}$$

7.2.2 Dérivées premières centrées

Formules à trois points

$$u'(x_j) \simeq \frac{U_{j+1} - U_{j-1}}{2h} + \mathcal{O}(h^2) \tag{7.53}$$

Formules à 5 points

$$u'(x_j) \simeq \frac{-U_{j+2} + 8U_{j+1} - 8U_{j-1} + U_{j-1}}{12h} + \mathcal{O}(h^4) \tag{7.54}$$

7.2.3 Dérivées secondes décentrées

Formules à trois points

$$\begin{aligned} u''(x_j) &\simeq \frac{U_{j+2} - 2U_{j+1} + U_j}{h^2} + \mathcal{O}(h) \\ u''(x_j) &\simeq \frac{U_j - 2U_{j-1} + U_{j-2}}{h^2} + \mathcal{O}(h) \end{aligned} \tag{7.55}$$

Formules à cinq points

$$\begin{aligned} u''(x_j) &\simeq \frac{-U_{j+3} + 4U_{j+2} - 5U_{j+1} + 2U_j}{h^2} + \mathcal{O}(h^2) \\ u''(x_j) &\simeq \frac{2U_j - 5U_{j-1} + 4U_{j-2} - U_{j-3}}{h^2} + \mathcal{O}(h^2) \end{aligned} \tag{7.56}$$

7.2.4 Dérivées secondes centrées

Formules à trois points

$$u''(x_j) \simeq \frac{U_{j+1} - 2U_j + U_{j-1}}{h^2} + \mathcal{O}(h^2) \tag{7.57}$$

Formules à cinq points

$$u''(x_j) \simeq \frac{-U_{j+2} + 16U_{j+1} - 30U_j + 16U_{j-1} - U_{j-2}}{12h^2} + \mathcal{O}(h^4) \tag{7.58}$$

7.3 Résolution d'un problème différentiel

7.3.1 Problème continu

Considérons le problème suivant :

$$\begin{aligned} &\mathcal{L}(u) = f \quad \text{sur} \quad]x_a, x_b[\\ &u(x_a) = u_a \\ &u(x_b) = u_b \end{aligned} \tag{7.59}$$

où $\mathcal{L}$ est défini dans la section 7.2. Ce problème est appelé *problème continu* et la solution exacte est notée $u(x)$.

7.3.2 Problème discret

Pour écrire le problème discret nous allons, après avoir maillé le segment $[x_a, x_b]$, remplacer les dérivées continues par leurs approximations en fonction des valeurs discrètes dans l'équation à résoudre pour les points intérieurs au domaine. Nous serons conduits

à la résolution d'un système linéaire dont la structure dépendra du nombre de points choisis pour discrétiser les dérivées. Le nombre minimal est de trois points (pour avoir une approximation de la dérivée seconde) et nous obtiendrons alors une matrice tridiagonale. Elle présente l'avantage de pouvoir être stockée en mémoire à l'aide de trois vecteurs de taille N (au lieu d'une matrice N × N) et qu'il existe des méthodes de résolution tenant compte de cette structure, diminuant ainsi considérablement le temps de calcul.

Maillage

Soit N le nombre de points que l'on se donne *a priori* sur l'intervalle $[x_a; x_b]$. On définit le pas h supposé constant. Tout point i de cet intervalle, sera défini par son abscisse $x_i = x_a + (i-1)h$, avec $h = \frac{x_b - x_a}{\mathrm{N}-1}$ comme indiqué sur la figure 7.2.

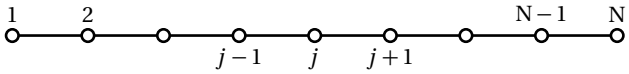


figure 7.2 - *Maillage du problème discret monodimensionnel*

Équation discrète

Point courant (intérieur)

Au point j, d'abscisse x_j, nous écrivons l'équation $\mathscr{L}(u) = f$ en remplaçant les dérivées première et seconde par les formules :

$$\begin{aligned} u'(x_j) &\simeq \frac{\theta \mathrm{U}_{j+1} + (1-2\theta)\mathrm{U}_j - (1-\theta)\mathrm{U}_{j-1}}{h} \\ u''(x_j) &\simeq \frac{\mathrm{U}_{j+1} - 2\mathrm{U}_j + \mathrm{U}_{j-1}}{h^2} \end{aligned} \tag{7.60}$$

soit, en remplaçant dans l'équation initiale :

$$\frac{a(x_j)}{h^2}(\mathrm{U}_{j+1} - 2\mathrm{U}_j + \mathrm{U}_{j-1}) + \frac{b(x_j)}{h}(\theta \mathrm{U}_{j+1} + (1-2\theta)\mathrm{U}_j - (1-\theta)\mathrm{U}_{j-1}) + c(x_j)\mathrm{U}_j = f(x_j) \tag{7.61}$$

On note $a_j = a(x_j)$, $b_j = b(x_j)$, $c_j = c(x_j)$ et $f_j = f(x_j)$. En ordonnant en fonction de U_{j-1}, U_j et U_{j+1}, il vient :

$$\mathrm{U}_{j+1}\left(\frac{a_j}{h^2} + \theta\frac{b_j}{h}\right) + \mathrm{U}_j\left(\frac{-2a_j}{h^2} + (1-2\theta)\frac{b_j}{h} + c_j\right) + \mathrm{U}_{j-1}\left(\frac{a_j}{h^2} - (1-\theta)\frac{b_j}{h}\right) = f_j \tag{7.62}$$

soit encore :

$$r_j^+ \mathrm{U}_{j+1} + r_j^c \mathrm{U}_j + r_j^- \mathrm{U}_{j-1} = d_j \tag{7.63}$$

avec :

$$r_j^+ = \frac{a_j}{h^2} + \theta\frac{b_j}{h}; \quad r_j^c = -2\frac{a_j}{h^2} + (1-2\theta)\frac{b_j}{h} + c_j; \quad r_j^- = \frac{a_j}{h^2} - (1-\theta)\frac{b_j}{h}; \quad d_j = f_j$$

Nous allons écrire cette équation discrète pour chacun des N − 2 points intérieurs, soit $j = 2, \mathrm{N}-1$ et aurons donc à résoudre un système de N − 2 équations à N inconnues.

Pour fermer le système, il faut utiliser les conditions limites et écrire les deux équations discrètes correspondantes. Le système d'équations linéaires à résoudre s'écrit :

$$\begin{bmatrix} \mathrm{CL} & & & & & & & \\ r_2^- & r_2^c & r_2^+ & & & & & \\ & r_3^- & r_3^c & r_3^+ & & & & \\ & & & \ddots & & & & \\ & & & r_j^- & r_j^c & r_j^+ & & \\ & & & & & \ddots & & \\ & & & & & r_{\mathrm{N}-1}^- & r_{\mathrm{N}-1}^c & r_{\mathrm{N}-1}^+ \\ & & & & & & & \mathrm{CL} \end{bmatrix} \begin{pmatrix} \mathrm{U}_1 \\ \mathrm{U}_2 \\ \mathrm{U}_3 \\ \vdots \\ \mathrm{U}_j \\ \vdots \\ \mathrm{U}_{\mathrm{N}-1} \\ \mathrm{U}_{\mathrm{N}} \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_j \\ \vdots \\ d_{\mathrm{N}-1} \\ d_{\mathrm{N}} \end{pmatrix} \tag{7.64}$$

On remarque que le système ainsi obtenu est un système tridiagonal ; pour conserver ces propriétés, il convient d'écrire les conditions limites sur deux points au plus de discrétisation.

Conditions limites (Dirichlet)

Dans le cas où nous avons des conditions de Dirichlet (la valeur de la fonction u est donnée aux points x_a et x_b), nous pourrons écrire :

– au premier point 1 :

$$r_1^c \mathrm{U}_1 + r_1^+ \mathrm{U}_2 = d_1 \tag{7.65}$$

soit $r_1^c = 1$, $r_1^+ = 0$ et $d_1 = u(x_a) = u_a$;

– au dernier point N :

$$r_{\mathrm{N}}^- \mathrm{U}_{\mathrm{N}-1} + r_{\mathrm{N}}^c \mathrm{U}_{\mathrm{N}} = d_{\mathrm{N}} \tag{7.66}$$

soit $r_{\mathrm{N}}^- = 0$, $r_{\mathrm{N}}^c = 1$ et $d_{\mathrm{N}} = u(x_b) = u_b$

et nous avons à résoudre un système linéaire $\mathrm{N} \times \mathrm{N}$.

Conditions limites (Neumann)

Approximation en deux points : dans le cas où nous avons des conditions de Neumann (la valeur de la dérivée u' est donnée aux points x_a et x_b), nous allons utiliser au point 1 (ou N) une formule discrète d'approximation en deux points de la dérivée première :

$$u'(x_1) \simeq \frac{\mathrm{U}_2 - \mathrm{U}_1}{h} \quad \text{ou} \quad u'(x_{\mathrm{N}}) \simeq \frac{\mathrm{U}_{\mathrm{N}} - \mathrm{U}_{\mathrm{N}-1}}{h} \tag{7.67}$$

Au point 1, par identification avec la formule générale $r_1^c \mathrm{U}_1 + r_1^+ \mathrm{U}_2 = d_1$, nous obtenons $r_1^c = -1/h$, $r_1^+ = 1/h$ et $d_1 = u'(x_a) = u'_a$. Le point N nous conduit à $r_{\mathrm{N}}^c = 1/h$, $r_1^+ = -1/h$ et $d_{\mathrm{N}} = u'(x_b) = u'_b$.

Approximation en trois points : si on veut avoir plus de précision sur l'expression de la dérivée première, il est possible d'écrire une formule en trois points (1, 2 et 3) ou (N − 2, N − 1 et N), dans ce cas, la matrice $\mathrm{N} \times \mathrm{N}$ n'est plus tridiagonale et pour retrouver cette structure, il faut réduire l'ordre de ce système de $2(\mathrm{N}-2) \times (\mathrm{N}-2)$. Pour

cela, il faut éliminer U_1 entre les 2 premières équations et U_{N-1} entre les deux dernières. D'une façon générale, si on considère les deux premières équations, nous aurons :

$$\begin{array}{l} r_1^c U_1 + r_1^+ U_2 + r_1^{++} U_3 = d_1 \\ r_2^- U_1 + r_2^c U_2 + r_2^+ U_3 = d_2 \end{array} \tag{7.68}$$

la première équation fournit la valeur de U_1 en fonction de U_2 et U_3, soit :

$$U_1 = \frac{1}{r_1^c}(d_1 - r_1^+ U_2 + r_1^{++} U_3) \tag{7.69}$$

et en reportant dans la deuxième équation, il vient :

$$(r_2^c - r_1^+ r_2^- r_1^c)U_2 + (r_2^+ - r_1^{++} r_2^- r_3^c)U_3 = d_2 - r_2^- r_1^c d_1 \tag{7.70}$$

que l'on peut écrire sous la forme :

$$r_2'^c U_2 + r_2'^+ U_3 = d_2' \tag{7.71}$$

avec :

$$r_2'^c = r_2^c - \frac{r_1^+ r_2^-}{r_1^c} \qquad r_2'^+ = r_2^+ - \frac{r_1^{++} r_2^-}{r_3^c} \qquad d_2' = d_2 - \frac{r_2^- d_1}{r_1^c}$$

Approximation en trois points avec un point fictif : afin de préserver le caractère tridiagonal de la matrice et d'éviter une première élimination de U_1 entre les équations au point 2 et l'écriture de la condition limite, on peut introduire un point fictif noté U_0 et introduire la condition limite dans l'équation différentielle discrète écrite au point 1. Le calcul de $u'(x)$ au point 1 en utilisant les points 2 et 0 donne :

$$u'(x_1) = \frac{U_2 - U_0}{2h} \tag{7.72}$$

l'équation au point 1 s'écrit $r_1^- U_0 + r_1^c U_1 + r_1^+ U_2 = d_1$ et en remplaçant U_0 par son expression tirée de l'équation précédente $U_0 = U_2 - 2hu_a'$, soit :

$$r_1^c U_1 + (r_1^+ + r_1^-)U_2 = d_1 + 2hr_1^- u_a' \tag{7.73}$$

Le système final à résoudre sera donc de la forme suivante :

$$\begin{bmatrix} r_1^c & r_1^+ & & & & & & \\ r_2^- & r_2^c & r_2^+ & & & & & \\ & r_3^- & r_3^c & r_3^+ & & & & \\ & & & \ddots & & & & \\ & & & r_j^- & r_j^c & r_j^+ & & \\ & & & & & \ddots & & \\ & & & & & r_{N-1}^- & r_{N-1}^c & r_{N-1}^+ \\ & & & & & & r_N^- & r_N^c \end{bmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ \vdots \\ U_j \\ \vdots \\ U_{N-1} \\ U_N \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_j \\ \vdots \\ d_{N-1} \\ d_N \end{pmatrix} \tag{7.74}$$

Résolution

Pour résoudre le problème discrétisé suivant :

$$r_j^+ \mathrm{U}_{j+1} + r_j^c \mathrm{U}_j + r_j^- \mathrm{U}_{j-1} = d_j \tag{7.75}$$

qui correspond à un système de N équations linéaires à N inconnues, on utilise souvent l'algorithme de Thomas ou dit de double balayage. Cet algorithme correspond en fait à une élimination de Gauss appliquée à un système tridiagonal. On pose la relation de récurrence :

$$\mathrm{U}_j = \mathrm{E}_j \mathrm{U}_{j+1} + \mathrm{F}_j, \quad j = \mathrm{N}-1, 1 \tag{7.76}$$

ce qui permet, connaissant la valeur de $\mathrm{U_N}$, les valeurs des E_j et des F_j, de calculer la solution. Au point N, nous avons :

$$r_\mathrm{N}^c \mathrm{U_N} + r_\mathrm{N}^- \mathrm{U}_{\mathrm{N}-1} = d_\mathrm{N} \tag{7.77}$$

avec $r_\mathrm{N}^+ = 0$, quel que soit le type de la condition limite. Nous obtenons les relations de récurrence suivantes :

$$\mathrm{E}_j = -\frac{r_j^+}{r_j^- \mathrm{E}_{j-1} + r_j^c} \quad \text{et} \quad \mathrm{F}_j = \frac{d_j - r_j^- \mathrm{F}_{j-1}}{r_j^- \mathrm{E}_{j-1} + r_j^c}, \quad j = 2, \mathrm{N} \tag{7.78}$$

ces deux relations permettent de calculer, par récurrence, les valeurs des coefficients E_j et F_j, connaissant les valeurs de E_1 et de F_1 qui sont obtenues par identification, à partir de la première équation $r_1^c \mathrm{U}_1 + r_1^+ \mathrm{U}_2 = d_1$ soit $\mathrm{U}_1 = -r_1^+ r_1^c \mathrm{U}_2 + d_1 r_1^c$ ce qui donne :

$$\mathrm{E}_1 = -\frac{r_1^+}{r_1^c} \quad \text{et} \quad \mathrm{F}_1 = \frac{d_1}{r_1^c} \tag{7.79}$$

Algorithme de Thomas

L'algorithme est alors le suivant, et ce quel que soit le type de conditions limites utilisées :

- *Premier balayage :*

$$\begin{array}{lll} j = 1 & \mathrm{E}_1 = -\dfrac{r_1^+}{r_1^c} & \mathrm{F}_1 = \dfrac{d_1}{r_1^c} \\ j = 2, \mathrm{N} & \mathrm{E}_j = -\dfrac{r_j^+}{r_j^- \mathrm{E}_{j-1} + r_j^c} & \mathrm{F}_j = \dfrac{d_j - r_j^- \mathrm{F}_{j-1}}{r_j^- \mathrm{E}_{j-1} + r_j^c} \end{array} \tag{7.80}$$

- *Deuxième balayage :*

$$\begin{array}{ll} j = \mathrm{N} & \mathrm{U_N} = \mathrm{F_N} \\ j = \mathrm{N}-1, 1 & \mathrm{U}_j = \mathrm{E}_j \mathrm{U}_{j+1} + \mathrm{F}_j \end{array} \tag{7.81}$$

En fait, si on prend soin de définir les vecteurs E_j et F_j de 0 à N et le vecteur solution U_j de 0 à $\mathrm{N}+1$, l'algorithme précédent est beaucoup plus simple car on peut supprimer les formules de *bord.*

- *Premier balayage :*

$$\begin{array}{ll} & \mathrm{E_0} = \mathrm{F_0} = 0, \quad r_1^- = 0, \quad r_\mathrm{N}^+ = 0 \\ j = 1, \mathrm{N} & \mathrm{E}_j = -\dfrac{r_j^+}{r_j^- \mathrm{E}_{j-1} + r_j^c} \quad \mathrm{F}_j = \dfrac{d_j - r_j^- \mathrm{F}_{j-1}}{r_j^- \mathrm{E}_{j-1} + r_j^c} \end{array} \tag{7.82}$$

- *Deuxième balayage :*

$$\begin{aligned} &\mathrm{U}_{\mathrm{N}+1}=0 \quad (\mathrm{E}_{\mathrm{N}} \text{ est nul car } r_{\mathrm{N}}^{+}=0) \\ j=\mathrm{N},1 \quad &\mathrm{U}_j=\mathrm{E}_j\mathrm{U}_{j+1}+\mathrm{F}_j \end{aligned} \tag{7.83}$$

Erreur de troncature

Considérons le problème discrétisé suivant :

$$r_j^{+}\mathrm{U}_{j+1}+r_j^{c}\mathrm{U}_j+r_j^{-}\mathrm{U}_{j-1}=d_j \tag{7.84}$$

Si on remplace les valeurs discrètes U_j par les valeurs continues correspondantes $u(x_j)$, et si nous effectuons le développement en séries de Taylor de l'expression ainsi obtenue, nous devons retrouver le problème continu plus des termes du premier ordre en h (ou $\mathscr{O}(h)$) au moins. Dans ce cas, le schéma global sera consistant. Si d'autres termes $\mathscr{O}(1)$ apparaissent, le schéma ne sera pas consistant, c'est à dire que l'on traitera en fait un autre problème continu. En remplaçant les valeurs discrètes par les valeurs continues, nous aurons :

$$r_j^{+}u(x_{j+1})+r_j^{c}u(x_j)+r_j^{-}u(x_{j-1})-d_j=0 \tag{7.85}$$

et en utilisant le développement en séries de Taylor, nous obtenons :

$$\begin{aligned} 1 \quad &\equiv d_j=f_j \\ u(x_j) \quad &\equiv r_j^{+}+r_j^{c}+r_j^{-} \\ u'(x_j) \quad &\equiv h(r_j^{+}-r_j^{-}) \\ u''(x_j) \quad &\equiv \frac{h^2}{2!}(r_j^{+}+r_j^{-}) \\ u^{(3)}(x_j) \quad &\equiv \frac{h^3}{3!}(r_j^{+}-r_j^{-}) \\ u^{(p)}(x_j) \quad &\equiv \frac{h^p}{p!}(r_j^{+}+(-1)^p r_j^{-}) \end{aligned}$$

soit :

$$\begin{aligned} u(x_j)(r_j^{+}+r_j^{c}+r_j^{-})+hu'(x_j)(r_j^{+}-r_j^{-})+\frac{h^2}{2!}u''(x_j)(r_j^{+}+r_j^{-}) \\ +\frac{h^3}{3!}u'''(x_j)(r_j^{+}-r_j^{-})+\frac{h^4}{4!}u^{(4)}(r_j^{+}+r_j^{-})-f_j+\mathscr{O}(h^5)=0 \end{aligned} \tag{7.86}$$

En reprenant le schéma précédent, nous avons :

$$r_j^{+}+r_j^{c}+r_j^{-}=c_j; \qquad r_j^{+}-r_j^{-}=\frac{b_j}{h}; \qquad r_j^{+}+r_j^{-}=\frac{2a_j}{h^2}-\frac{(1-2\theta)b_j}{h} \tag{7.87}$$

et en remplaçant dans l'expression précédente :

$$\begin{aligned} \left(a_j u''(x_j)+b_j u(x_j)+c_j u(x_j)-f_j\right)-(1-2\theta)\frac{hb_j}{2!}u''(x_j)+ \\ \frac{h^2 b_j}{3!}u'''(x_j)+\frac{h^2 a_j}{4!}u^{(4)}(x_j)=\mathscr{O}(h^5) \end{aligned} \tag{7.88}$$

On remarque que le terme entre crochets est nul car c'est le problème différentiel initial et que le premier terme de l'erreur de troncature est :

$$-(1-2\theta)\frac{hb_j}{2!}u''(x_j) \tag{7.89}$$

Si $\theta = 1/2$, ce premier terme de l'erreur de troncature est nul et l'erreur est bien alors en $\mathcal{O}(h^2)$; le schéma est d'ordre 2, sinon, elle est $\mathcal{O}(h)$ seulement et le schéma sera du premier ordre.

7.4 Admissibilité d'un schéma

Dans ce paragraphe, nous allons montrer, sur un exemple simple que le problème discret peut dans certaines conditions conduire à des solutions non admissibles, autrement dit, que la solution du problème discret n'est pas compatible avec la solution du problème continu.

7.4.1 Problème continu d'advection-diffusion

Considérons le problème continu suivant :

$$\begin{aligned} &\mathcal{L}(u) = 0 \quad \text{sur} \quad]0,\,1[\\ &u(0) = 1 \\ &u(1) = 0 \end{aligned} \tag{7.90}$$

avec $\mathcal{L}(u) = a\,u'' + b\,u'$, où a et b sont des constantes. Par rapport au modèle proposé, on remarque que le coefficient c (terme $c\,u$) est nul ; le terme $(c\,u)$ est un terme source, dont l'influence sera traitée plus loin.

Ce problème admet une solution analytique de la forme :

$$u(x) = \alpha e^{-bx/a} + \beta \tag{7.91}$$

où α et β sont des coefficients déterminés de façon à vérifier les conditions limites. En effet, d'une façon générale, si l'opérateur $\mathcal{L}(u)$ est défini par $\mathcal{L}(u) = a\,u'' + b\,u' + c\,u$ et si les coefficients a, b et c sont constants, la solution analytique est $u(x) = \mathrm{A}e^{q_1 x} + \mathrm{B}e^{q_2 x}$, où q_1 et q_2 sont les deux racines du polynôme caractéristique $aq^2 + bq + c = 0$ et où A et B sont des coefficients déterminés de façon à vérifier les deux conditions limites du problème. Ici, les deux racines sont respectivement $q_1 = -b/a$ et $q_2 = 0$. La solution du problème s'écrit alors (si $b \neq 0$) :

$$u(x) = \frac{e^{-bx/a} - e^{-b/a}}{1 - e^{-b/a}} \tag{7.92}$$

solution exacte du problème continu. Dans le cas où $b = 0$, la solution du problème est la droite d'équation :

$$u(x) = 1 - x \tag{7.93}$$

Lorsque b/a prend des valeurs élevées (en valeur absolue), nous avons des solutions de type couche limite. Mathématiquement parlant, si a est petit, le problème est de type perturbation singulière. Plus simplement, si a et b sont positifs, on constate sur la figure 7.5, que la solution $u(x)$ décroît très fortement pour des valeurs de x inférieures à 0,1 et qu'à partir des valeurs supérieures à 0,4, la solution varie faiblement.

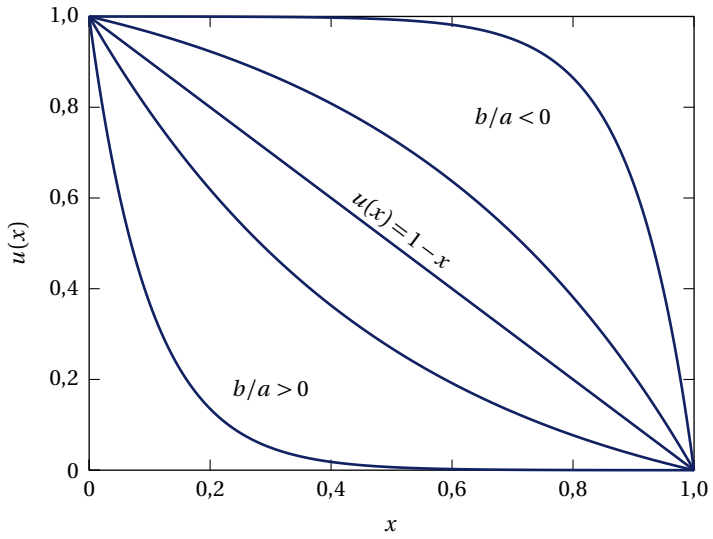


figure 7.3 - *Solution exacte* (7.92) *du problème continu* (7.90)

7.4.2 Problème discret

Après discrétisation, le problème peut se mettre sous la forme suivante :

$$r_j^+ \mathrm{U}_{j+1} + r_j^c \mathrm{U}_j + r_j^- \mathrm{U}_{j-1} = 0, \quad i = 2, \mathrm{N}-1 \tag{7.94}$$

Comme les coefficients a, b et c sont constants, les coefficients r sont indépendants de x, donc de j d'où la relation :

$$r^+ \mathrm{U}_{j+1} + r^c \mathrm{U}_j + r^- \mathrm{U}_{j-1} = 0 \tag{7.95}$$

et comme conditions limites $\mathrm{U}_1 = u(0) = 1$ et $\mathrm{U}_{\mathrm{N}-1} = u(1) = 0$, avec :

$$r^+ = \frac{a}{h^2} + \frac{\theta b}{h}; \quad r^c = -\frac{2a}{h^2} + (1-2\theta)\frac{b}{h}; \quad r_j^- = \frac{a}{h^2} - (1-\theta)\frac{b}{h}; \quad d = 0$$

Dans ce cas particulier (coefficients r constants), le problème discret admet lui aussi une solution analytique qui se met sous la forme :

$$\mathrm{U}_j = \gamma q_1^{j-1} + \delta q_2^{j-1} \tag{7.96}$$

où q_1 et q_2 sont les 2 racines du polynôme caractéristique $r^+ q^2 + r^c q + r^- = 0$. On utilise pour ce faire le théorème dit des suites récurrentes. On peut vérifier facilement que $r^+ + r^c + r^- = 0$ (car $c = 0$) et que l'une des racines est donc égale à 1 et que l'autre est égale à r^-/r^+ ; on notera $q_1 = r^-/r^+$ et $q_2 = 1$. Si la racine est double ($q_1 = q_2 = 1$) ; nous aurons une solution particulière de la solution correspondant au cas où $b = 0$.

Dans cet exemple, nous avons donc à la fois la solution analytique du problème continu et celle du problème discret ; cela va donc nous permettre de comparer ces solutions de façon détaillée. Avant de déterminer la valeur des coefficients γ et δ on peut tout de suite voir que si q_1 est négatif, alors q_1^{j-1} présentera des oscillations paires-impaires et que cette partie de la solution ne peut pas représenter une solution monotone, si $|q_1| < 1$, ces oscillations seront atténuées, mais elles seront amplifiées si $|q_1| > 1$. Si on calcule les

coefficients γ et δ de façon à satisfaire les conditions limites, la solution exacte du problème discret s'écrit :

$$\mathrm{U}_j = \frac{q_1^{j-1} - q_1^{\mathrm{N}-1}}{1 - q_1^{\mathrm{N}-1}} = \frac{\left(\frac{r^-}{r^+}\right)^{j-1} - \left(\frac{r^-}{r^+}\right)^{\mathrm{N}-1}}{1 - \left(\frac{r^-}{r^+}\right)^{\mathrm{N}-1}} \tag{7.97}$$

	problème continu	problème discret
polynôme caractéristique	$aq^2+bq+c=0$	$r^+q^2+r^cq+r^-=0$
solution générale	$u(x)=ae^{q_1x}+be^{q_2x}$	$\mathrm{U}_j=\gamma q_1^{j-1}+\delta q_2^{j-1}$
q_1	$c=-b/a$	$c=r^-/r^+$
q_2	0	1
solution du problème	$u(x)=\frac{e^{cx}-e^c}{1-e^c}$	$\mathrm{U}_j=\frac{c^{j-1}-c^{\mathrm{N}-1}}{1-c^{\mathrm{N}-1}}$

tableau 7.1 - *Comparaison problème continu/problème discret*

7.4.3 Comparaison des solutions

Nous allons donc pouvoir comparer les solutions des deux problèmes, continu et discret, car les solutions sont ici connues analytiquement. Ces deux solutions sont respectivement :

$$u(x) = \frac{e^{-bx/a} - e^{-b/a}}{1 - e^{-b/a}} \quad \text{et} \quad \mathrm{U}_j = \frac{q_1^{j-1} - q_1^{\mathrm{N}-1}}{1 - q_1^{\mathrm{N}-1}} \tag{7.98}$$

Afin de simplifier l'analyse, on supposera que $a > 0$:

1. si $b > 0$ alors $e^{-b/a} < 1$, la fonction $e^{-bx/a}$ sera décroissante et $u(x)$ sera aussi décroissante ;
2. si $b < 0$ alors $e^{-b/a} > 1$, la fonction $e^{-bx/a}$ sera croissante et $u(x)$ sera donc décroissante.

Pour avoir des solutions discrètes qui ont le même comportement que les solutions continues nous devrons donc avoir

1. si $b > 0$ alors $0 < q_1 < 1$;
2. si $b < 0$ alors $q_1 > 1$.

En utilisant les résultats du paragraphe précédent, nous pouvons écrire :

$$\frac{r^-}{r^+} = \frac{1 + (\theta - 1)bh/a}{1 + \theta bh/a} \tag{7.99}$$

où θ est le paramètre de décentration et h le pas de discrétisation ; en posant $\rho = bh/a$ nous avons

$$q_1 = \frac{r^-}{r^+} = \frac{1 + (\theta - 1)\rho}{1 + \theta\rho} \tag{7.100}$$

Nous allons donc examiner les diverses valeurs de θ, à savoir 0, 1/2 ou 1 et écrire les conditions que doit vérifier ρ pour que la solution discrète soit admissible. Le nombre

sans dimension ρ, est appelé nombre de Reynolds de maille, ou nombre de Péclet de maille.

Par analogie avec la mécanique des fluides, le nombre sans dimension bh/a est souvent appelé nombre de Reynolds de maille car il est de la même forme que le nombre de Reynolds classique défini par $\mathrm{Re} = \mathrm{UL}/\nu$ où U est une vitesse, L une longueur et ν la viscosité. De même, si on fait référence aux problèmes thermiques, on définira le nombre de Péclet de maille $\mathrm{P}_e = \mathrm{UL}/\kappa$ où U est une vitesse, L une longueur et κ la diffusivité thermique.

1. $\theta = 1 : q_1 = \frac{1}{1+\rho}$
   la condition $q_1 > 0$ implique $\rho \in]-1, \infty[$:
   – si $b > 0$ alors $\rho > 0$ et $q_1 < 1$ donc $\rho \in]0, \infty[$ et aucune limitation sur h ;
   – si $b < 0$ alors $\rho < 0$ et $q_1 > 1$, d'où $\rho \in]-1, 0[$.
   Dans ce cas, on voit que le choix $\theta = 1$ est intéressant si $b > 0$ car il n'y a pas de limite sur la valeur de ρ, donc du pas h.

2. $\theta = 0 : q_1 = 1 - \rho$
   la condition $q_1 > 0$ implique $\rho \in]-\infty, 1[$:
   – si $b > 0$ alors $\rho > 0$ et $q_1 < 1$, d'où $\rho \in]0, 1[$;
   – si $b < 0$ alors $\rho < 0$ et $q_1 > 1$ donc $\rho \in]-\infty, 0[$ et aucune limitation sur h.
   Dans ce cas, on voit que le choix $\theta = 0$ est intéressant si $b < 0$ car il n'y a pas de limite sur la valeur de ρ, donc du pas h.

3. $\theta = 1/2 : q_1 = \frac{1-\rho/2}{1+\rho/2}$ et la condition $q_1 > 0$ implique $\rho \in]-2, 2[$:
   – si $b > 0$ alors $\rho > 0$ et $q_1 < 1$, ce qui implique $\rho \in]0, 2[$;
   – si $b < 0$ alors $\rho < 0$ et $q_1 > 1$, ce qui implique $\rho \in]-2, 0[$.

Le tableau 7.2 récapitule les conditions à satisfaire pour que le schéma discret soit admissible.

	$\theta = 0$	$\theta = 1/2$	$\theta = 1$
$b < 0$	$\rho \in]-\infty, 0[$	$\rho \in]-2, 0[$	$\rho \in]-1, 0[$
$b < 0$	$\rho \in]0, 1[$	$\rho \in]0, 2[$	$\rho \in]0, \infty[$

tableau 7.2 - *Conditions d'admissibilité*

Finalement, rappelons les résultats principaux :

(i) pour un schéma centré ($\theta = 1/2$), le pas de discrétisation doit vérifier $|bh/a| < 2$, soit encore $h < 2|a/b|$. Par conséquent, un schéma d'ordre 2 nécessite la vérification d'une condition pour obtenir une solution numérique admissible ;

(ii) pour un schéma décentré ($\theta = 0$ ou $\theta = 1$), la précision est $\mathcal{O}(h)$ et paramètre de décentration, peut être choisi de façon à ce que les solutions discrètes soient admissibles. Dans ce cas, le pas de discrétisation sera déterminé uniquement sur un critère de précision de la solution.

7.4.4 Décentration « upwind »

Dans de nombreux ouvrages, la notion de décentration « upwind » ou « au vent » est utilisée. Cette notion, contraire de « sous le vent » ou « downwind », est familière aux voileux mais il convient de l'expliquer. Nous allons considérer un exemple en mécanique des fluides où les équations à résoudre sont souvent des équations de transport et sont par exemple (pour la température T) de la forme suivante :

$$\mathrm{V}\frac{\mathrm{dT}}{\mathrm{d}x} = k\frac{\mathrm{d}^2\mathrm{T}}{\mathrm{d}x^2} \tag{7.101}$$

où V est une vitesse (de convection ou advection) et k un coefficient de diffusion. Cette équation peut se mettre sous la forme $\mathscr{L}(u) = a\,u'' + b\,u'$, avec $a = k$ et $b = -\mathrm{V}$:

- si $\mathrm{V} > 0$ alors $b < 0$ et on a intérêt à choisir $\theta = 0$ pour calculer la dérivée première de T, soit :

$$\frac{\mathrm{dT}}{\mathrm{d}x} \simeq \frac{\mathrm{T}_j - \mathrm{T}_{j-1}}{h} \tag{7.102}$$

  donc on choisit le point $j-1$ pour calculer la dérivée, comme V est positif, on choisit un point *au vent*, c'est-à-dire en remontant la vitesse.
- si $\mathrm{V} < 0$ alors $b > 0$ et on a intérêt à choisir $\theta = 1$ pour calculer la dérivée première de T, soit :

$$\frac{\mathrm{dT}}{\mathrm{d}x} \simeq \frac{\mathrm{T}_{j+1} - \mathrm{T}_j}{h} \tag{7.103}$$

  donc on choisit le point $j+1$ pour calculer la dérivée, comme V est négatif, la encore, le point choisi est un point "au vent", c'est-à-dire en remontant la vitesse.

Cette notion de décentration "upwind" est utilisée dans de nombreux domaines de la physique, même si la notion de vitesse est absente.

figure 7.4 - *Notion de décentration upwind :* × *points utilisés pour le calcul de la dérivée première en* j *,* ∘ *point non utilisé*

Nous avons choisi comme exemple l'équation de la chaleur où l'inconnue est la température car c'est un scalaire. En mécanique des fluides, on a aussi des équations de transport qui traduisent la conservation de la quantité de mouvement, mais celles-ci sont couplées et non-linéaires, ainsi dans le cas d'un écoulement bidimensionnel, nous avons :

$$\begin{aligned}
\mathrm{U}\frac{\partial \mathrm{U}}{\partial x} + \mathrm{V}\frac{\partial \mathrm{U}}{\partial y} &= -\frac{1}{\rho}\frac{\partial \mathrm{P}}{\partial x} + \nu\left(\frac{\partial^2 \mathrm{U}}{\partial x^2} + \frac{\partial^2 \mathrm{U}}{\partial x^2}\right) \\
\mathrm{U}\frac{\partial \mathrm{V}}{\partial x} + \mathrm{V}\frac{\partial \mathrm{V}}{\partial y} &= -\frac{1}{\rho}\frac{\partial \mathrm{P}}{\partial y} + \nu\left(\frac{\partial^2 \mathrm{U}}{\partial x^2} + \frac{\partial^2 \mathrm{U}}{\partial x^2}\right)
\end{aligned} \tag{7.104}$$

7.4.5 Admissibilité et dominance diagonale

La possibilité d'utiliser certaines méthodes pour la résolution numérique de systèmes linéaires est souvent liée à la satisfaction d'une condition de *dominance diagonale* qui

stipule :

$$|a_{ii}| = \sum_{j=1, j\neq i}^{\mathrm{N}} |a_{ij}| \quad \forall i = 1, \mathrm{N} \tag{7.105}$$

soit dans le cas d'une matrice tridiagonale :

$$|r_j^c| \geqslant |r_j^+| + |r_j^-| \quad \forall i = 1, \mathrm{N} \tag{7.106}$$

Dans l'exemple traité, $a\,u'' + b\,u' = 0$, les valeurs des coefficients r sont :

$$\begin{aligned} r^+ &= \frac{a}{h^2} + \frac{\theta b}{h} = \frac{a}{h^2}\left(1 + \theta\frac{bh}{a}\right) = \frac{a}{h^2}(1+\theta\rho) \\ r^- &= \frac{a}{h^2} + \frac{(\theta-1)b}{h} = \frac{a}{h^2}\left(1 + (\theta-1)\frac{bh}{a}\right) = \frac{a}{h^2}(1+(\theta-1)\rho) \\ r^c &= \frac{-2a}{h^2} + (1-2\theta)\frac{b}{h} = \frac{a}{h^2}\left(-2 + (1-2\theta)\frac{bh}{a}\right) = \frac{a}{h^2}(-2+(1-2\theta)\rho) \end{aligned} \tag{7.107}$$

Nous sommes donc ramenés à étudier en fonction de θ et ρ les expressions $(1+\theta\rho)$ puis $1+(\theta-1)\rho$ et enfin $-2+(1-2\theta)\rho$:

(i) $\theta = 1$: dans ces conditions, $r_j^+ = 1+\rho$, $r_j^- = 1$ et $r_j^c = -(2+\rho)$ et la condition de dominance diagonale impose $\rho \in]-1, \infty[$;

(ii) $\theta = 0$: dans ces conditions, $r_j^+ = 1$, $r_j^- = 1-\rho$ et $r_j^c = -(2+\rho)$ et la condition de dominance diagonale impose $\rho \in]-\infty, 1[$;

(iii) $\theta = 1/2$ pour ce schéma centré, $r_j^+ = 1+\rho/2$, $r_j^- = 1-\rho/2$ et $r_j^c = -2$ et la condition de dominance diagonale impose $\rho \in]-2, 2[$.

On constate donc, que lorsqu'il n'y a pas dominance diagonale, nous avons des solutions oscillantes, donc non admissibles ; par contre, il faut vérifier que la solution discrète a même sens de variation que la solution continue.

7.4.6 Problème continu de diffusion avec terme source

Considérons le problème continu suivant :

$$\begin{aligned} &\mathscr{L}(u) = 0 \quad \text{sur} \quad]0,1[\\ &u(0) = 1 \\ &u'(1) = 0 \end{aligned} \tag{7.108}$$

avec $\mathscr{L}(u) = a\,u'' + c\,u$, où a et c sont des constantes (on supposera que a est positif et c négatif). Par rapport au modèle proposé, on remarque que le coefficient b (terme $b\,u'$) est nul, donc il n'y a pas d'advection. Par contre, le terme source $(c\,u)$ n'est pas nul. Ce problème modèle correspond à l'étude de la répartition de température dans une ailette de largeur unité et de section A constante. L'équation (7.109) représente ce phénomène :

$$\frac{\mathrm{d}^2\mathrm{T}}{\mathrm{d}x^2} + \frac{h\mathrm{S}}{\kappa\mathrm{A}}(\mathrm{T}_\infty - \mathrm{T}) = 0 \tag{7.109}$$

où T est la température de l'ailette, T_∞ la température extérieure, κ, le coefficient de diffusion du matériau constituant l'ailette, S, le périmètre de la section de l'ailette, A,

son aire et h, le coefficient d'échange avec l'extérieur. On appelle T_0 la température de l'ailette en $x = 0$. En absence d'échange avec l'extérieur, si l'extrémité droite est isolée, la température dans l'ailette est constante est égale à T_0. Si $h \neq 0$, la solution analytique du problème physique est :

$$T(x) = T_\infty + (T_0 - T_\infty)\frac{ch[\alpha(L-x)]}{ch(\alpha L)} \tag{7.110}$$

avec $\alpha^2 = \frac{hS}{A\kappa}$. Dans le cas de l'équation modèle donnée au début de ce paragraphe, la solution analytique sera :

$$U(x) = \frac{ch[\alpha(1-x)]}{ch(\alpha)} \tag{7.111}$$

avec $\alpha^2 = -\frac{c}{a}$ (donc c est ici négatif) :

$$r^+ U_{j+1} + r^c U_j + r^- U_{j-1} = d \tag{7.112}$$

avec : $r^+ = a/h^2$, $r^c = -2a/h^2 + c$, $r^- = a/h^2$ et $d = 0$. Compte tenu des valeurs des coefficients a et c, on constate que les deux coefficients r^+ et r^- sont positifs, alors que le coefficient r^c est lui négatif ; dans ce cas, on vérifie aisément que la matrice du système linéaire discret est à dominance diagonale. Comme dans l'exemple précédent, les coefficients r sont constants et le problème discret admet lui aussi une solution analytique qui se met sous la forme :

$$U_j = \gamma q_1^{j-1} + \delta q_2^{j-1} \tag{7.113}$$

où q_1 et q_2 sont les 2 racines du polynôme caractéristique $r^+q^2 + r^c q + r^- = 0$. Ces racines se calculent classiquement et $\Delta = (r^c)^2 - 4r^+r^- = c^2 + \frac{4ac}{h^2}$. Comme c est ici négatif, les deux racines sont réelles et on vérifie que le produit de ces racines est égal à 1 ($r^+ = r^-$). La somme de ces racines étant positive les deux racines sont positives et $\gamma = \frac{1}{\delta}$.

7.5 Applications

7.5.1 Étude de l'admissibilité

On considère le problème d'advection diffusion sur l'intervalle [0 10], avec comme valeurs des paramètres du problème $a = 1$ et $b = 10$. On étudie l'influence du paramètre de décentration sur l'admissibilité du schéma. Le tableau 7.3 donne les caractéristiques des valeurs intervenant dans cette étude. Nous sommes dans le cas où b est positif, donc q_1 doit être positif et inférieur à 1. Les calculs ont été effectués en simple précision et la figure 7.5 représente les solutions obtenues avec $h = 0{,}125$, pour chaque type de dérivée choisie, à savoir décentrées (avant ou arrière) ou centrées. La notion avant ou arrière fait référence à la droite ou la gauche par rapport au point considéré.

7.5.2 Étude de la précision

Dans ce paragraphe, on s'intéresse à la précision des schémas. Dans les cas considérés, on a choisi $a = b = 1$ et un nombre de points variant de 11 à 10 001. Les calculs sont effectués à l'aide des deux schémas décentrés et du schéma centré. En double précision, on

nb de points	h	θ	ρ	q_1
11	1	0 0,5 1	10	−9 −0,666667 0,00909094
21	0,5	0 0,5 1	5	−4 −0,428574 0,1666667
41	0,25	0 0,5 1	2,5	−1,5 −0,111111 0,2857143
81	0,125	0 0,5 1	1,25	−0,25 0,2307692 0,4444444

tableau 7.3 - *Valeurs pour l'étude d'admissibilité des schémas*

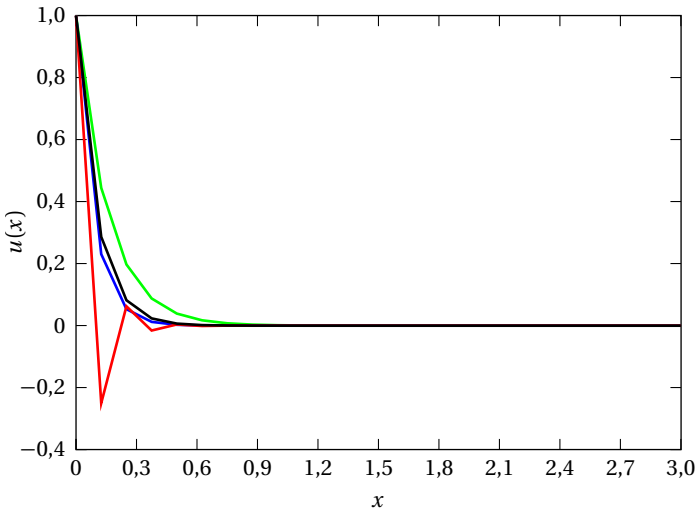


figure 7.5 - *Solutions exacte* (——) *et approchées* (7.98), $bh/a = 1{,}25$; $\theta = 0$ (——), $bh/a = 1{,}25$; $\theta = 0{,}5$ (——) *et* $bh/a = 1{,}25$; $\theta = 1$ (——) *du problème continu* (7.90)

remarque que l'erreur maximale obtenue à l'aide des deux schémas décentrés ($\mathcal{O}(h)$) varie linéairement en fonction du pas ; la pente de cette droite est bien -1. Pour le schéma centré, la pente est égale à -2, ce qui indique que le schéma est en $\mathcal{O}(h^2)$. Par contre, on remarque que pour de petites valeurs du pas, il y a de fortes oscillations de l'erreur maximale ; cela indique que la précision de la machine est incompatible avec la précision théorique.

La figure 7.6 donne les évolutions des erreurs maximales en fonction du pas h.

7.6 Coordonnées cylindriques ou sphériques

7.6.1 Équations monodimensionnelles

On considère maintenant un problème différentiel d'ordre 2 d'une fonction à une variable r correspondant à l'écriture d'un Laplacien en coordonnées cylindriques ou sphériques sur un intervalle $[r_a, r_b]$. Ce type de problème correspond à l'écriture de l'équation de la chaleur dans un cylindre ou dans une sphère, pour des problèmes axisymétriques ou à symétrie sphérique. Dans ce paragraphe, comme dans les précédents, nous ne

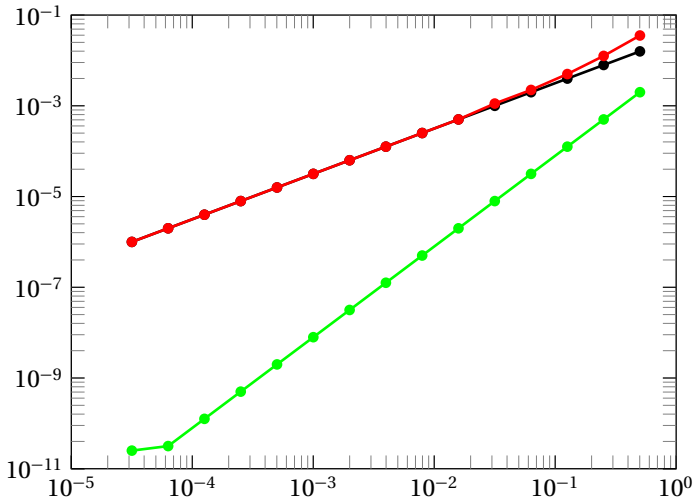


figure 7.6 - *Influence de la décentration :* $\theta = 1$ (●—●), $\theta = 0{,}5$ (●—●) *et* $\theta = 0$ (●—●)

considérons que des problèmes monodimensionnels. Les équations correspondantes s'écrivent :

$$\frac{d^2T}{dr^2} + \frac{m}{r}\frac{dT}{dr} = 0 \tag{7.114}$$

avec $m = 1$ en coordonnées cylindriques et $m = 2$ en coordonnées sphériques. Ces équations ne sont pas valables en $r = 0$, car il y a une singularité apparente en $(1/r)$; il faut lever l'indétermination, en effet, en $r = 0$, pour des raisons de symétrie, $\frac{dT}{dr} = 0$. En utilisant en $r = 0$ la règle de Lhospital, on obtient alors l'équation suivante :

pas h

$$(m+1)\frac{d^2T}{dr^2} = 0 \tag{7.115}$$

avec $m = 1$ en coordonnées cylindriques $m = 2$ en coordonnées sphériques

Rappel de la règle de Lhospital en $r = 0$:

$$\frac{1}{r}\frac{dT}{dr} = \frac{\frac{d}{dr}\left(\frac{dT}{dr}\right)}{\frac{d}{dr}(r)} = \frac{d^2T}{dr^2} \tag{7.116}$$

7.6.2 Conditions limites

Deux cas distincts sont à considérer, selon que le domaine d'étude contient $r = 0$ ou non. Dans le cas où $r_a = 0$, la condition limite $\frac{dT}{dr} = 0$ est incontournable car elle est une hypothèse du calcul. On peut imposer une condition supplémentaire sous réserve d'avoir unicité de la solution du problème. En effet, si on impose en $r = r_b$ une condition de Neumann, le problème (Neumann Neumann) admet une infinité de solutions et on peut fixer alors la température au centre du domaine. Par contre, en utilisant la condition de Neumann $\frac{dT}{dr} = 0$ ou condition de symétrie, on peut évaluer la dérivée seconde en $r = 0$, par :

$$\frac{d^2T}{dr^2} \simeq \frac{2(T_2 - T_1)}{\Delta r^2} \tag{7.117}$$

où Δr est le pas d'espace. En effet, nous avons :

$$\frac{\mathrm{d}^2\mathrm{T}}{\mathrm{d}r^2} \simeq \frac{\mathrm{T}_2 - 2\mathrm{T}_1 + \mathrm{T}_0}{\Delta r^2} \tag{7.118}$$

et :

$$\frac{\mathrm{dT}}{\mathrm{d}r} \simeq \frac{\mathrm{T}_2 - \mathrm{T}_0}{2\Delta r} = 0 \tag{7.119}$$

Dans le cas où $r_a \neq 0$, les mêmes problèmes que ceux exposés dans les paragraphes précédents subsistent.

8

Méthodes mehrstellen

8.1 Construction du schéma

8.1.1 Problème

Soit une fonction u supposée continue et dérivable sur un intervalle $[a, b]$ et un opérateur différentiel $\mathscr{L}$ tel que $\mathscr{L}(u) = a\,u'' + b\,u' + c\,u$ où les coefficients a, b et c sont des fonctions de la variable x. On pose le problème suivant :

$$\begin{aligned} &\mathscr{L}(u) = 0 \qquad \text{pour} \quad x \in]x_a,\, x_b[\\ &u(x_a) = u_a \\ &u(x_b) = u_b \end{aligned} \tag{8.1}$$

8.1.2 Principe

Au lieu de se donner des formules de discrétisation permettant de remplacer chaque dérivée dans l'équation $\mathscr{L}(u) = f$, on va se donner à priori une équation aux différences et déterminer les coefficients de telle sorte que le problème discret approche le problème continu avec la meilleure précision possible. On pose :

$$r_j^+ \mathrm{U}_{j+1} + r_j^c \mathrm{U}_j + r_j^- \mathrm{U}_{j-1} = h^2 \left(q_j^+ f_{j+1} + q_j^c f_j + q_j^- f_{j-1} \right) \tag{8.2}$$

où f_{j+1}, f_j et f_{j-1}, sont les valeurs du second membre aux points $j+1$, j et $j-1$. Par rapport à ce qui a été vu précédemment, la valeur du second membre apparaît en trois points au lieu d'un seul ; on peut donc prévoir que la formule ainsi obtenue sera plus précise. Il y a donc 6 coefficients à déterminer r_j^+, r_j^c, r_j^-, q_j^+, q_j^c et q_j^-, en fait, ces coefficients ne sont pas indépendants et il ne faudra qu'en déterminer cinq.

Pour déterminer ces coefficients, nous allons écrire le développement en séries de Taylor de la formule (8.2) et le mettre sous la forme $\mathrm{E} = \sum_{k=0}^{\infty} \mathrm{T}_k\, u^{(k)}$ et annuler les premiers termes de ce développement. Dans ce développement, le terme f_{j+1} est remplacé par le développement de $\mathscr{L}(u)$ au point x_{j+1}, soit :

$$f_{j+1} = a(x_{j+1})u''(x_{j+1}) + b(x_{j+1})u'(x_{j+1}) + c(x_{j+1}) \tag{8.3}$$

8.1.3 Coefficients constants

Afin d'alléger les calculs, on supposera que les coefficients a, b et c sont indépendants de x ; les formules correspondantes au cas où a, b et c sont des fonctions de x seront données au paragraphe suivant. On rappelle les développements en séries de Taylor qui

sont utilisés, où h est le pas de discrétisation supposé constant :

$$\begin{aligned} u(x_{j-1}) &= u(x_j) - hu'(x_j) + \frac{h^2}{2!}u''(x_j) + \ldots + (-1)^k \frac{h^k}{k!}u^{(k)}(x_j) + \mathrm{O}(h^{k+1}) \\ u(x_{j-1}) &= u(x_j) + hu'(x_j) + \frac{h^2}{2!}u''(x_j) + \ldots + \frac{h^k}{k!}u^{(k)}(x_j) + \mathrm{O}(h^{k+1}) \end{aligned} \tag{8.4}$$

de même, nous aurons, pour les dérivées première et seconde :

$$\begin{aligned} u'(x_{j-1}) &= u'(x_j) - hu''(x_j) + \frac{h^2}{2!}u'''(x_j) + \ldots + (-1)^k \frac{h^k}{k!}u^{(k+1)}(x_j) + \mathrm{O}(h^{k+1}) \\ u''(x_{j-1}) &= u''(x_j) - hu'''(x_j) + \frac{h^2}{2!}u^{(4)}(x_j) + \ldots + (-1)^k \frac{h^k}{k!}u^{(k+2)}(x_j) + \mathrm{O}(h^{k+1}) \end{aligned} \tag{8.5}$$

et :

$$\begin{aligned} u'(x_{j+1}) &= u'(x_j) + hu''(x_j) + \frac{h^2}{2!}u'''(x_j) + \ldots + \frac{h^k}{k!}u^{(k+1)}(x_j) + \mathrm{O}(h^{k+1}) \\ u''(x_{j+1}) &= u''(x_j) + hu'''(x_j) + \frac{h^2}{2!}u^{(4)}(x_j) + \ldots + \frac{h^k}{k!}u^{(k+2)}(x_j) + \mathrm{O}(h^{k+1}) \end{aligned} \tag{8.6}$$

Calculons les coefficients T_k. Tout d'abord :

$$\mathrm{T}_0 = \left(r_j^+ + r_j^c + r_j^-\right) - ch^2\left(q_j^+ + q_j^c + q_j^-\right) \tag{8.7}$$

les coefficients r_j proviennent du développement du membre de droite de l'équation aux différences et les coefficients q_j du second membre.

$$\begin{aligned} \mathrm{T}_1 &= h\left(r_j^+ - r_j^-\right) - \left(ch^3\left(q_j^+ - q_j^-\right) + bh^2\left(q_j^+ + q_j^c + q_j^-\right)\right) \\ \mathrm{T}_2 &= \frac{h^2}{2!}\left(r_j^+ + r_j^-\right) - \left(c\frac{h^4}{2!}\left(q_j^+ - q_j^-\right) + bh^3\left(q_j^+ - q_j^-\right) + ah^2\left(q_j^+ + q_j^c + q_j^-\right)\right) \\ \mathrm{T}_3 &= \frac{h^3}{3!}\left(r_j^+ - r_j^-\right) - \left(c\frac{h^5}{3!}\left(q_j^+ - q_j^-\right) + b\frac{h^4}{2!}\left(q_j^+ + q_j^-\right) + \frac{h^3}{3!}\left(q_j^+ - q_j^-\right)\right) \end{aligned} \tag{8.8}$$

de façon générale, on peut obtenir :

$$\begin{aligned} \mathrm{T}_k = \frac{h^k}{k!}\left(r_j^+ + (-1)^k r_j^-\right) - \Bigg(& c\frac{h^{k+2}}{k!}\left(q_j^+ + (-1)^k q_j^-\right) + b\frac{h^{k+1}}{(k-1)!}\left(q_j^+ + (-1)^{k+1} q_j^-\right) \\ & + a\frac{h^k}{(k-2)!}\left(q_j^+ + (-1)^k q_j^-\right)\Bigg) \end{aligned} \tag{8.9}$$

Écrivons que les trois premiers coefficients T_0, T_1 et T_2 sont nuls ; ce faisant, le schéma de discrétisation obtenu sera consistant avec l'équation différentielle à résoudre. Nous pouvons ainsi exprimer les coefficients r_j en fonction des coefficients q_j :

$$\begin{aligned} r_j^+ + r_j^c + r_j^- &= ch^2(q_j^+ + q_j^c + q_j^-) \\ r_j^+ - r_j^- &= bh(q_j^+ + q_j^c + q_j^-) + ch^2(q_j^+ - q_j^-) \\ r_j^+ + r_j^- &= 2a(q_j^+ + q_j^c + q_j^-) + 2bh(q_j^+ - q_j^-) + ch^2(q_j^+ + q_j^-) \end{aligned} \tag{8.10}$$

Ces relations permettent de calculer les coefficients r_j^+, r_j^c et r_j^- en fonction des coefficients q_j^+, q_j^c et q_j^-, soit :

$$\begin{aligned} r_j^+ &= a(q_j^+ + q_j^c + q_j^-) + bh^2(3q_j^+ + q_j^c - q_j^-) + ch^2 q_j^+ \\ r_j^- &= a(q_j^+ + q_j^c + q_j^-) + bh^2(q_j^+ - q_j^c - 3q_j^-) + ch^2 q_j^+ \\ r_j^c &= ch^2(q_j^+ + q_j^c + q_j^- - r_j^+ - r_j^-) = ch^2 q_j^c - 2a(q_j^+ + q_j^c q_j^-) - 2bh(q_j^+ - q_j^-) \end{aligned} \tag{8.11}$$

où encore en ordonnant par rapport aux coefficients q :

$$\begin{aligned} r_j^+ &= q_j^+\left(a+\frac{3bh}{2}+ch^2\right)+q_j^c\left(a+\frac{bh}{2}\right)+q_j^-\left(a-\frac{bh}{2}\right)\\ r_j^- &= q_j^+\left(a+\frac{bh}{2}\right)+q_j^c\left(a-\frac{bh}{2}\right)+q_j^-\left(a-\frac{3bh}{2}+ch^2\right)\\ r_j^c &= q_j^+\left(-2a-\frac{2b}{h}\right)+q_j^c\left(ch^2-2a\right)+q_j^-\left(-2a-\frac{3bh}{2}\right) \end{aligned} \tag{8.12}$$

Si on choisit $q_j^+ = q_j^- = 0$ et $q_j^c = 1$, on peut calculer les coefficients r_j, et on obtient :

$$r_j^+ + r_j^c + r_j^- = ch^2; \qquad r_j^+ - r_j^- = bh; \qquad r_j^+ + r_j^- = 2a \tag{8.13}$$

soit encore :

$$r_j^+ = a+\frac{bh}{2}; \qquad r_j^- = a-\frac{bh}{2}; \qquad r_j^c = -2a+ch^2 \tag{8.14}$$

en reportant dans l'équation de discrétisation, il vient :

$$\left(a+\frac{bh}{2}\right)\mathrm{U}_{j+1}+\left(-2a+ch^2\right)\mathrm{U}_j+\left(a-\frac{bh}{2}\right)\mathrm{U}_{j-1} = h^2 f_j \tag{8.15}$$

et en divisant par h^2, nous avons :

$$\left(\frac{a}{h^2}+\frac{b}{2h}\right)\mathrm{U}_{j+1}+\left(-\frac{2a}{h^2}+c\right)\mathrm{U}_j+\left(\frac{a}{h^2}-\frac{b}{2h}\right)\mathrm{U}_{j-1} = f_j \tag{8.16}$$

soit encore :

$$a\frac{\mathrm{U}_{j+1}-2\mathrm{U}_j+\mathrm{U}_{j-1}}{h^2}+b\frac{\mathrm{U}_{j+1}-\mathrm{U}_{j-1}}{2h}+c\mathrm{U}_j = f_j \tag{8.17}$$

ce qui correspond à la formule obtenue avec la méthode des différences finies classiques en utilisant une formule centrée pour la dérivée première. À ce stade, on constate que seule la formulation centrée est obtenue, en effet, si on discrétise la dérivée première à l'aide d'une formule décentrée, l'erreur est en $\mathrm{O}(h)$ et en u'', terme qui a été supposé nul dans la construction précédente. Pour obtenir les formulations décentrées, il ne faut plus annuler le terme contenant la dérivée seconde mais écrire qu'il est en $\mathrm{O}(h)$, soit :

$$r_j^+ + r_j^c + r_j^- = ch^2; \qquad r_j^+ - r_j^- = bh; \qquad r_j^+ + r_j^- = 2a+2\alpha h \tag{8.18}$$

On obtient les valeurs suivantes des coefficients r :

$$r_j^+ = a+\alpha h+\frac{bh}{2}; \qquad r_j^- = a+\alpha h-\frac{bh}{2}; \qquad r_j^c = -2a-2\alpha h+ch^2 \tag{8.19}$$

et l'équation de discrétisation devient :

$$a\frac{\mathrm{U}_{j+1}-2\mathrm{U}_j+\mathrm{U}_{j-1}}{h^2}+\alpha\frac{\mathrm{U}_{j+1}-2\mathrm{U}_j+\mathrm{U}_{j-1}}{h}+b\frac{\mathrm{U}_{j+1}-\mathrm{U}_{j-1}}{2h}+c\mathrm{U}_j = f_j \tag{8.20}$$

et si on pose $\alpha = \alpha' b$, nous obtenons :

$$a\frac{\mathrm{U}_{j+1}-2\mathrm{U}_j+\mathrm{U}_{j-1}}{h^2}+b\frac{(1+2\alpha')\mathrm{U}_{j+1}-4\alpha'\mathrm{U}_j-(1-2\alpha')\mathrm{U}_{j-1}}{2h}+c\mathrm{U}_j = f_j \tag{8.21}$$

Le terme en facteur de b est bien l'expression d'une dérivée première décentrée, en effet, si $\alpha' = 0$, on retrouve l'expression de la dérivée centrée, si $\alpha' = 1/2$, on retrouve l'expression de la dérivée décentrée à gauche et enfin si $\alpha' = -1/2$ l'expression de la dérivée à droite. Dans un précédent chapitre, la formule générale de la dérivée première en trois points est donnée sous la forme :

$$u'(x_j) \simeq \frac{\theta U_{j+1} + (1-2\theta)U_j - (1-\theta)U_{j-1}}{h} \tag{8.22}$$

ce qui donne la correspondance entre α' et θ, à savoir $1+2\alpha' = 2\theta$, soit $\alpha' = \theta - 1/2$.

Déterminons les coefficients q_j à l'aide des coefficients T_3 et T_4. Comme l'équation aux différences est homogène, nous pourrons déterminer deux coefficients indépendants. Écrivons que T_3 et T_4 sont nuls. Il vient :

$$\begin{aligned} &q_j^+(6a+2bh) + q_j^-(-6a+2bh) = bhq_j^c \\ &q_j^+(5a+bh) + q_j^-(5a-bh) = aq_j^c \end{aligned} \tag{8.23}$$

Le déterminant de ce système est $\mathrm{Det} = 60a^2 - 4b^2h^2$ et les deux solutions sont :

$$\begin{aligned} q_j^+ &= \frac{q_j^c}{\mathrm{Det}}(6a^2 + 3abh - b^2h^2) \\ q_j^- &= \frac{q_j^c}{\mathrm{Det}}(6a^2 - 3abh - b^2h^2) \end{aligned} \tag{8.24}$$

On a considéré le cas où le déterminant est non nul ; celui-ci s'annule pour la valeur particulière $h = \pm 15a/b$ où encore $bh/a = \pm 15$. En choisissant la valeur de q_j^c égale au déterminant Det (supposé non nul), nous aurons alors :

$$\begin{aligned} q_j^+ &= 6a^2 + 3abh - b^2h^2 \\ q_j^c &= 60a^2 - 4b^2h^2 \\ q_j^- &= 6a^2 - 3abh - b^2h^2 \\ r_j^+ &= 72a^3 + 36a^2bh - 3b^3h^3 + c(6a^2h^2 + 3abh^3 - b^2h^4) \\ r_j^c &= -144a^3 + c(60a^2h^2 - 4b^2h^4) \\ r_j^- &= 72a^3 - 36a^2bh + 3b^3h^3 + c(6a^2h^2 - 3abh^3 - b^2h^4) \end{aligned} \tag{8.25}$$

On remarque que les deux coefficients q_j^+ et q_j^- sont symétriques et que l'on passe de l'un à l'autre en changeant h en $-h$.

Dans le cas où le déterminant est nul, nous allons vérifier que le problème a une solution, pour cela, il suffit de réécrire les équations sous la forme :

$$\begin{aligned} &q_j^+(6a+2bh) - bhq_j^c = -q_j^-(-6a+2bh) \\ &q_j^+(5a+bh) - aq_j^c = -q_j^-(5a-bh) \end{aligned} \tag{8.26}$$

Dans ces conditions, le déterminant Det′ est égal à $Det' = -6a^2 + 3bh + b^2h^2$ qui est non nul pour toute valeur de h. On peut remarquer que $\mathrm{Det}' = q_j^-$ et que l'on retrouve les mêmes valeurs que dans le cas précédent.

Calculons les coefficients T_5 et T_6 :

$$\begin{aligned} T_5 &= \frac{h^5}{5!}\left(r_j^+ - r_j^-\right) - \left(c\frac{h^7}{5!}\left(q_j^+ - q_j^-\right) + b\frac{h^6}{4!}\left(q_j^+ + q_j^-\right) + a\frac{h^5}{3!}\left(q_j^+ - q_j^-\right)\right) \\ T_6 &= \frac{h^6}{6!}\left(r_j^+ + r_j^-\right) - \left(c\frac{h^8}{6!}\left(q_j^+ + q_j^-\right) + b\frac{h^7}{5!}\left(q_j^+ - q_j^-\right) + a\frac{h^6}{4!}\left(q_j^+ + q_j^-\right)\right) \end{aligned} \tag{8.27}$$

on peut vérifier facilement que $r_j^+ - r_j^-$ est en O(h), tout comme $q_j^+ - q_j^-$. Le terme $r_j^+ + r_j^-$ est en O(1) de même que le terme $q_j^+ + q_j^-$ et par conséquent les termes négligés sont en O(h^6) à la fois pour T_5 et T_6 ; l'erreur de troncature de ce schéma sera donc en O(h^4).

Explicitons l'erreur de troncature :

$$\begin{aligned}
E &= \sum_{k=0}^{N} T_k u^{(k)} \\
&= \left(r_j^+ U_{j+1} + r_j^c U_j + r_j^- U_{j-1}\right) - ch^2\left(q_j^+ + q_j^c + q_j^-\right) \\
&= h^2\left(q_j^+ + q_j^c + q_j^-\right)\left(au'' + bu' + cu - f + h^4\left(-\frac{b}{5!}u^{(5)} - \frac{2a}{6!}u^{(6)}\right)\right) \\
&= h^2\left(q_j^+ + q_j^c + q_j^-\right)\left(au'' + bu' + cu - f + O(h^4)\right)
\end{aligned} \tag{8.28}$$

8.1.4 Coefficients variables

Si maintenant on considère que les coefficients a, b et c sont dépendants de x, les indices vont intervenir dans l'écriture des développements de l'opérateur $l(u)$. Dans ces conditions, les formules du paragraphe précédent s'écriront :

$$\begin{aligned}
q_j^+ &= 6a_j a_{j-1} + h\left(5a_{j-1}b_j - 2a_j b_{j-1}\right) - b_j b_{j-1} h^2 \\
q_j^c &= 60a_{j-1}a_{j+1} + 16h\left(a_{j-1}b_{j+1} - a_{j+1}b_{j-1}\right) - 4h^2 b_{j-1}b_{j+1} \\
q_j^- &= 6a_j a_{j+1} + h\left(2a_j b_{j+1} - 5a_j b_{j+1}\right) - b_j b_{j+1} h^2 \\
r_j^+ &= \frac{q_j^-\left(2a_{j-1} - hb_{j-1}\right) + q_j^c\left(2a_j + hb_j\right) + q_j^+\left(2a_{j+1} + 3hb_{j+1} + 2h^2 c_{j+1}\right)}{2} \\
r_j^- &= \frac{q_j^-\left(2a_{j-1} - 3hb_{j-1} + 2h^2 c_{j-1}\right) + q_j^c\left(2a_j + hb_j\right) + q_j^+\left(2a_{j+1} + hb_{j+1}\right)}{2} \\
r_j^c &= -\left(r_j^+ + r_j^-\right) + h^2\left(q_j^- c_{j-1} + q_j^c c_j + q_j^+ c_{j+1}\right)
\end{aligned} \tag{8.29}$$

On vérifie que :

$$\begin{aligned}
r_j^+ + r_j^- &= q_j^-(2a_{j-1} - 2hb_{j-1} + h^2 c_{j-1}) + q_j^+(2a_{j+1} + 2hb_{j+1} + h^2 c_{j+1}) \\
r_j^c &= -q_j^-(2a_{j-1} - 2hb_{j-1}) - q_j^c(2a_j - h^2 c_j) - q_j^+(2a_{j+1} + 2hb_{j+1})
\end{aligned} \tag{8.30}$$

8.2 Propriétés du schéma

8.2.1 Application : formule de Numerov

On considère les coefficients $[a,b,c] = [1,0,0]$. L'opérateur $\mathscr{L}(u)$ s'écrit donc u''. En remplaçant a, b et c par ces valeurs dans les formules suivantes :

$$\begin{aligned}
q_j^+ &= 6a^2 + 3abh - b^2h^2 \\
q_j^c &= 60a^2 - 4b^2h^2 \\
q_j^+ &= 6a^2 - 3abh - b^2h^2 \\
r_j^+ &= 72a^3 + 36a^2bh - 3b^3h^3 + 6ach^2 + 3abch^3 - b^2ch^4 \\
r_j^c &= -144a^3 + 60a^2h^2 - 2b^2h^4 \\
r_j^- &= 72a^3 - 36a^2bh + 3b^3h^3 + 6ach^2 - 3abch^3 - b^2ch^4
\end{aligned} \tag{8.31}$$

on obtient $q_j^+ = 6$, $q_j^c = 60$ et $q_j^- = 6$ d'une part et $r_j^+ = 72$, $r_j^c = -144$ et $q_j^- = 72$ d'autre part. Si on appelle U''_{j+1}, U''_j et U''_{j-1} les valeurs des dérivées secondes aux points

$j-1$, j et $j+1$, en remplaçant les seconds membres f_{j+1}, f_j et f_{j-1} par U''_{j+1}, U''_j et U''_{j-1} dans l'équation générale, nous obtenons une relation entre les valeurs discrètes en trois points d'une fonction et de sa dérivée seconde. Avec une telle formule, connaissant les valeurs discrètes d'une fonction en tout point avec suffisamment de précision, si on connaît deux conditions limites portant sur la dérivée seconde, on pourra calculer avec une précision $\mathrm{O}(h^4)$ les valeurs discrètes de la dérivée seconde :

$$72\mathrm{U}_{j+1}-144\mathrm{U}_j+72\mathrm{U}_{j-1}=h^2\left(6\mathrm{U}''_{j+1}+60\mathrm{U}''_j+6\mathrm{U}''_{j-1}\right) \tag{8.32}$$

soit encore :

$$\frac{\mathrm{U}_{j+1}-2\mathrm{U}_j+\mathrm{U}_{j-1}}{h^2}=\frac{\mathrm{U}''_{j+1}+10\mathrm{U}''_j+\mathrm{U}''_{j-1}}{12} \tag{8.33}$$

Il n'est pas possible d'utiliser le même raisonnement pour obtenir une formule similaire reliant les valeurs de la dérivée première et de la fonction en trois points ; pour cela, il faut utiliser des développements en série de Taylor directement. La formule de Numerov a d'ailleurs été obtenue de cette façon. En écrivant :

$$\frac{\mathrm{U}_{j+1}-2\mathrm{U}_j+\mathrm{U}_{j-1}}{h^2}=a\mathrm{U}''_{j+1}+b\mathrm{U}''_j+c\mathrm{U}''_{j-1} \tag{8.34}$$

et en développant U et U″ en séries de Taylor, on obtient :

$$\begin{aligned} u''(x_j)+\frac{2h^2}{4!}u^{(4)}(x_j)+\frac{2h^4}{6!}u^{(6)}(x_j)&=(a+b+c)u''(x_j)+(a-c)hu'''(x_j)\\ &+(a+c)\frac{h^2}{2!}u^{(4)}(x_j)+(a-c)\frac{h^3}{3!}u^{(5)}(x_j)+(a+c)\frac{h^4}{4!}u^{(6)}(x_j) \end{aligned} \tag{8.35}$$

nous pouvons donc choisir a, b et c de telle sorte que les termes en $\mathrm{O}(1)$ et en $\mathrm{O}(h^2)$ soient égaux dans cette équation, soit :

$$a+b+c=1;\qquad a-c=0;\qquad a+c=1/6 \tag{8.36}$$

ce qui donne $a=c=1/12$ et $b=10/12$; on retrouve ainsi la formule de Numerov et on vérifie bien que cette formule est en $\mathrm{O}(h^4)$. La formule reliant les valeurs de la fonction et de sa dérivée première s'obtient en écrivant :

$$\frac{\mathrm{U}_{j+1}-\mathrm{U}_{j-1}}{2h}=a\mathrm{U}'_{j+1}+b\mathrm{U}'_j+c\mathrm{U}'_{j-1} \tag{8.37}$$

et en calculant a, b et c. Le calcul montre que nous avons à satisfaire :

$$a+b+c=1;\qquad a-c=0;\qquad a+c=1/3 \tag{8.38}$$

d'où les valeurs $a=c=1/6$ et $c=2/3$ et la formule s'écrit alors :

$$\frac{\mathrm{U}_{j+1}-\mathrm{U}_{j-1}}{2h}=\frac{\mathrm{U}'_{j+1}+4\mathrm{U}'_j+\mathrm{U}'_{j-1}}{6} \tag{8.39}$$

8.2.2 Admissibilité

Reprenons l'exemple précédent où l'opérateur différentiel était de la forme $\mathscr{L}(u)=a\,u''+b\,u'$ avec a et b constants ; le problème était $\mathscr{L}(u)=a$ et on avait trouvé les conditions

d'admissibilité correspondant aux schémas centré et décentrés. Déterminons les conditions sur ρ ; en posant $\rho = bh/a$, nous avons $q_1 = \frac{r^-}{r^+}$ et $q_2 = 1$ car on vérifie que :

$$\begin{aligned} r_j^+ &= 72a^3 + 36a^2bh - 3b^3h^3 \\ r_j^c &= -144a3 \\ r_j^- &= 72a^3 - 36a^2bh + 3b^3h^3 \end{aligned} \tag{8.40}$$

on vérifie que la somme de ces coefficients est bien nulle et donc :

$$q_1 = \frac{r^-}{r^+} = \frac{24 - 12\rho + \rho^3}{24 + 12\rho - \rho^3} \tag{8.41}$$

1. si $b > 0$, $\rho > 0$ et nous devons avoir $q_1 \in]0,1[$ soit, $[24 - 12\rho + \rho^3 < 24 + 12\rho - \rho^3$ autrement dit $\rho(12 - \rho^2) > 0$ d'où la condition :

$$\rho \in]0, 2\sqrt{3}[\tag{8.42}$$

2. si $b < 0$, $\rho < 0$ et nous devons avoir $q_1 \in]1,\infty[$ soit, $24 - 12\rho + \rho^3 > 24 + 12\rho - \rho^3$ autrement dit $\rho(12 - \rho^2) < 0$ d'où la condition :

$$\rho \in]-2\sqrt{3}, 0[\tag{8.43}$$

En conclusion, on constate que, par rapport à un schéma centré d'ordre 2, non seulement la précision est augmentée (ordre 4), mais les conditions portant sur l'admissibilité sont moins strictes. ($2\sqrt{3}$ au lieu de 2).

8.2.3 Résultats

Le graphe 8.1 indique que l'erreur maximale entre solution numérique et solution exacte est bien d'ordre 4 du fait, dans le graphe log/log, de la présence d'une pente de -4. Par contre, très rapidement ($h > 0{,}001$) la précision de la machine est insuffisante et l'erreur augmente lorsque le pas diminue. Il est donc important de choisir un pas de discrétisation et une méthode compatibles avec la précision de la machine.

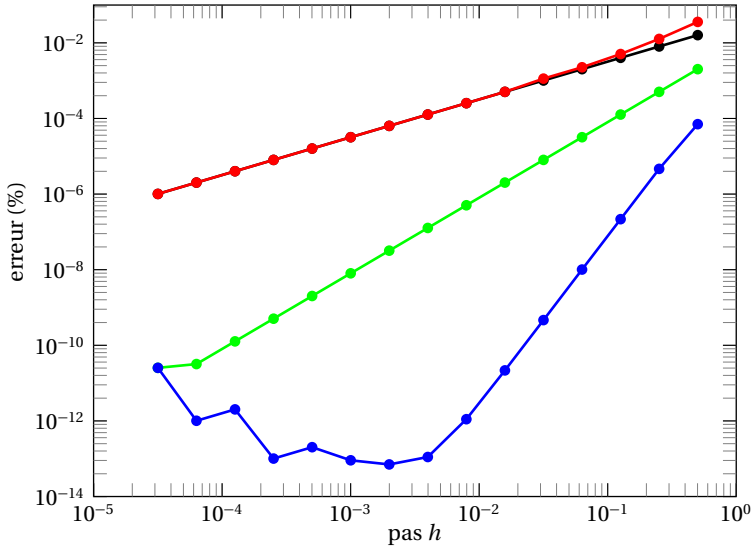


figure 8.1 - *Influence de la décentration :* $\theta = 1$ (●—●), $\theta = 0{,}5$ (●—●), $\theta = 0$ (●—●) *et mehr.* (●—●)

9

Méthodes hermitiennes

9.1 Principe général

9.1.1 Problème

Les méthodes hermitiennes sont des méthodes numériques utilisées pour résoudre des équations différentielles en prenant comme inconnues non seulement la valeur de la fonction en un point mais aussi celles de ses dérivées première et seconde. Ainsi, par point, nous aurons trois inconnues au lieu d'une et il faudra donc des équations supplémentaires pour fermer le système. Les inconnues étant au nombre de trois, au lieu d'avoir une matrice simple, nous aurons une matrice composée de blocs 3×3 ; une telle matrice est définie comme étant une matrice par blocs.

Le problème à résoudre est :

$$\begin{aligned} &\mathscr{L}(u) = a\,u'' + b\,u' + c\,u = f \quad \text{sur} \quad]x_a\ x_b[\\ &u(x_a) = u_a \\ &u(x_b) = u_b \end{aligned} \tag{9.1}$$

9.1.2 Équation

Si on écrit l'opérateur $\mathscr{L}$ au point x, d'abscisse $x = x_i$, nous aurons :

$$a(x_i)\mathrm{U}_i'' + b(x_i)\mathrm{U}_i' + c(x_i)\mathrm{U}_i = f(x_i) \tag{9.2}$$

où U_i, U_i' et U_i'' sont les valeurs approchées de la fonction $u(x)$ et de ses dérivées $u'(x)$ et $u''(x)$ en ce point i.

9.1.3 Construction des formules hermitiennes

On appelle formule hermitienne en trois points, une formule qui relie les valeurs d'une fonction et de ses dérivées premières et secondes sur ces trois points. En reprenant les notations de R. Peyret, nous pouvons écrire les formes générales suivantes, en considérant un pas constant h.

Fonction et dérivée première

Partons de l'expression suivante :

$$\mathrm{P}_i(\rho,\theta) = (\rho+\theta)\mathrm{U}'_{i+1} + 4\rho\,\mathrm{U}'_i + (\rho-\theta)\mathrm{U}'_{i-1} - \frac{(3\rho+2\theta)\mathrm{U}_{i+1} - 4\theta\,\mathrm{U}_i - (3\rho-2\theta)\mathrm{U}_{i-1}}{h} \tag{9.3}$$

On écrira :

$$\mathrm{P}_i(\rho,\theta) = 0 \tag{9.4}$$

Si on effectue un développement en séries de Taylor de cette formule, on montre que l'erreur de troncature est sous la forme :

$$\frac{h^3}{6}\theta u^{(4)} + \frac{h^4}{30}\rho u^{(5)} + \mathcal{O}(h^5) \tag{9.5}$$

dont les deux premiers termes ne peuvent pas être nuls simultanément. Ainsi, si θ est nul, nous aurons la formule :

$$\mathrm{P}_i(\rho,0) = \rho \mathrm{U}'_{i+1} + 4\rho \mathrm{U}'_i + \rho \mathrm{U}'_{i-1} - \frac{3\rho \mathrm{U}_{i+1} - 3\rho \mathrm{U}_{i-1}}{h} = 0 \tag{9.6}$$

ce qui conduit :

$$\frac{\mathrm{U}'_{i+1} + 4\mathrm{U}'_i + \mathrm{U}'_{i-1}}{6} = \frac{\mathrm{U}_{i+1} - \mathrm{U}_{i-1}}{2h} \tag{9.7}$$

dont la précision est d'ordre 4.

Si $\rho = 0$, $\mathrm{P}_i(0,\theta) = 0$ conduit à la relation :

$$\theta \mathrm{U}'_{i+1} - \theta \mathrm{U}'_{i-1} - \frac{2\theta \mathrm{U}_{i+1} - 4\theta \mathrm{U}_i + 2\theta \mathrm{U}_{i-1}}{h} = 0 \tag{9.8}$$

soit :

$$\frac{\mathrm{U}'_{i+1} - \mathrm{U}'_{i-1}}{2} = \frac{\mathrm{U}_{i+1} - 2\mathrm{U}_i + \mathrm{U}_{i-1}}{h} \tag{9.9}$$

Fonction, dérivée première et dérivée seconde

Partons maintenant de :

$$\begin{aligned} \mathrm{D}_i(\alpha,\beta,\gamma) = {} & (\alpha-\beta-\gamma)\mathrm{U}''_{i+1} + 4\alpha \mathrm{U}''_i + (\alpha+\beta-\gamma)\mathrm{U}''_{i-1} \\ & - \frac{(3\alpha-7\beta-5\gamma)\mathrm{U}'_{i+1} - 16\beta \mathrm{U}'_i - (3\alpha+7\beta-5\gamma)\mathrm{U}'_{i-1}}{h} \\ & - \frac{(15\beta+8\gamma)\mathrm{U}_{i+1} - 16\gamma \mathrm{U}_i - (15\beta-8\gamma)\mathrm{U}_{i-1}}{h^2} \end{aligned} \tag{9.10}$$

De même, on écrira :

$$\mathrm{D}_i(\alpha,\beta,\gamma) = 0 \tag{9.11}$$

L'erreur de troncature de cette formule est :

$$\frac{h^4}{90}(3\alpha - 2\gamma)u^{(6)} - \frac{h^5}{315}\beta u^{(7)} + \mathcal{O}(h^6) \tag{9.12}$$

En éliminant les dérivées premières, nous avons $\beta = 0$ et $3\alpha = 5\gamma$, d'où la formule, dite de Numerov :

$$\frac{\mathrm{U}''_{i+1} + 10\mathrm{U}''_i + \mathrm{U}''_{i-1}}{12} = \frac{\mathrm{U}_{i+1} - 2\mathrm{U}_i + \mathrm{U}_{i-1}}{h^2} \tag{9.13}$$

et la précision de cette formule est bien d'ordre 4.

D'autres formules peuvent être construites en utilisant d'autres points, par exemple i, $i+1$ et $i+2$, ou i, $i-1$ et $i-2$, et nous aurons des formules décentrées qui seront moins précises. Pour garder la même précision, il faut prendre en compte plusieurs points.

9.2 Résolution d'un problème

Nous allons donc résoudre le problème décrit au paragraphe 9.1.1 à l'aide de méthodes hermitiennes.

9.2.1 Système tridiagonal par blocs

En écrivant l'opérateur $\mathscr{L}(u)$ au point i et en utilisant les deux formules hermitiennes particulières définies ci-dessus, nous aurons :

$$\begin{aligned} &c_i \mathrm{U}_i + b_i \mathrm{U}'_i + a_i \mathrm{U}''_i = f_i \\ &\frac{\mathrm{U}'_{i+1} + 4\mathrm{U}'_i + \mathrm{U}'_{i-1}}{6} - \frac{\mathrm{U}_{i+1} - \mathrm{U}_{i-1}}{2h} = 0 \\ &\frac{\mathrm{U}''_{i+1} + 10\mathrm{U}''_i + \mathrm{U}''_{i-1}}{12} - \frac{\mathrm{U}_{i+1} - 2\mathrm{U}_i + \mathrm{U}_{i-1}}{h^2} = 0 \end{aligned} \tag{9.14}$$

soit sous forme matricielle :

$$\begin{bmatrix} 0 & 0 & 0 & c_i & b_i & a_i & 0 & 0 & 0 \\ \frac{1}{2h} & \frac{1}{6} & 0 & 0 & \frac{4}{6} & 0 & -\frac{1}{2h} & \frac{1}{6} & 0 \\ -\frac{1}{h^2} & 0 & \frac{1}{12} & \frac{2}{h^2} & 0 & \frac{10}{12} & -\frac{1}{h^2} & 0 & \frac{1}{12} \end{bmatrix} \begin{pmatrix} \mathrm{U}_{i-1} \\ \mathrm{U}'_{i-1} \\ \mathrm{U}''_{i-1} \\ \mathrm{U}_i \\ \mathrm{U}'_i \\ \mathrm{U}''_i \\ \mathrm{U}_{i+1} \\ \mathrm{U}'_{i+1} \\ \mathrm{U}''_{i+1} \end{pmatrix} = \begin{pmatrix} f_i \\ 0 \\ 0 \end{pmatrix} \tag{9.15}$$

En faisant de même pour les $\mathrm{N}-2$ points intérieurs ($i = 2, \ldots, \mathrm{N}-1$), on trouve une matrice tridiagonale par blocs 3×3. Pour fermer le système, il faut utiliser les conditions limites, or on ne connaît que les valeurs de la fonction aux deux extrémités. En utilisant des formules hermitiennes, on pourra trouver les quatre équations manquantes. Les méthodes de résolution de systèmes tridiagonaux par blocs sont analogues à celles utilisées pour la résolution de système tridiagonaux simples.

9.2.2 Système tridiagonal

Il est possible de se ramener à un système tridiagonal simple en considérant l'opérateur $\mathscr{L}(u)$ écrit en trois points $i-1$, i et $i+1$ et en utilisant des formules hermitiennes en nombre suffisant pour éliminer entre 7 équations les 6 dérivées inconnues. On aura donc une seule équation à 3 inconnues U_{i-1}, U_i et U_{i+1}. Cette méthode, proposée par Krause et Al.[1976] est en fait la méthode Mehrstellen qui peut être bâtie de façon différente. On considère les sept équations suivantes :

$$a_{i-1}\mathrm{U}''_{i-1} + b_{i-1}\mathrm{U}'_{i-1} + c_{i-1}\mathrm{U}_{i-1} = f_{i-1} \tag{9.16a}$$
$$a_i \mathrm{U}''_i + b_i \mathrm{U}'_i + c_i \mathrm{U}_i = f_i \tag{9.16b}$$
$$a_{i+1}\mathrm{U}''_{i+1} + b_{i+1}\mathrm{U}'_{i+1} + c_{i+1}\mathrm{U}_{i+1} = f_{i+1} \tag{9.16c}$$

$$D_i(0,-\gamma,\gamma)=0 \tag{9.16d}$$
$$D_i(\gamma,0,\gamma)=0 \tag{9.16e}$$
$$D_i(0,\gamma,\gamma)=0 \tag{9.16f}$$

$$P_i(\rho,0)=0 \tag{9.16g}$$

Les équations (9.16d), (9.16e) et (9.16f) permettent de calculer explicitement les dérivées secondes aux points $i-1$, i et $i+1$ en fonction des valeurs de la fonction et de la dérivée première.

En éliminant les 6 dérivées dans ces équations, on est conduit à une équation de la forme :

$$\begin{aligned}
&r_i^+ U_{i+1} + r_i^c U_i + r_i^- U_{i-1} = d_i \\
&U''_{i-1} = \frac{7U_{i+1} - 16U_i + 23U_{i-1}}{2h^2} - \frac{U'_{i+1} + 8U'_i + 6U'_{i-1}}{h} \\
&U''_i = \frac{-2U_{i+1} - 2U_{i-1}}{h^2} + \frac{U'_{i+1} - U'_{i-1}}{2h} \\
&U''_{i+1} = \frac{-23U_{i+1} + 16U_i - 7U_{i-1}}{2h^2} + \frac{6U'_{i+1} + 8U'_i - U'_{i-1}}{2h} \\
&\frac{U'_{i+1} + 4U'_i + U'_{i-1}}{6} = \frac{U_{i+1} - U_{i-1}}{2h}
\end{aligned} \tag{9.17}$$

Après calculs, nous avons, dans le cas ou a, b et c constants :

$$\begin{aligned}
r_i^+ &= \frac{72a^3 + 36a^2bh - 3b^3h^3 + c(6a^2h^2 + 3abh^3 - b^2h4)}{h^2} \\
r_i^c &= \frac{-144a^3 + c(60a^2h^2 - 2b^2h4)}{h^2} \\
r_i^- &= \frac{72a^3 - 36a^2bh + 3b^3h^3 + c(6a^2h^2 - 3abh^3 - b^2h^4)}{h^2} \\
d_i &= q_i^+ f_{i+1} + q_i^c f_i + q_i^- f_{i-1}
\end{aligned} \tag{9.18}$$

avec :

$$\begin{aligned}
q_i^+ &= 6a^2 - 3abh - b^2h^2 \\
q_i^c &= 60a^2 - 3b^2h^2 \\
q_i^- &= 6a^2 - 3abh - b^2h^2
\end{aligned} \tag{9.19}$$

10

Résolution de systèmes tridiagonal ou pentadiagonal

10.1 Méthode de double balayage

Un système tridiagonal peut se résoudre par la méthode de double balayage ou algorithme de Thomas. Cet algorithme doit être répété pour chaque second membre. Soit le système suivant à résoudre :

$$\begin{bmatrix} b_1 & c_1 & & & & & \\ a_2 & b_2 & c_2 & & & & \\ & & \ddots & & & & \\ & & a_i & b_i & c_i & & \\ & & & & \ddots & & \\ & & & & a_{n-1} & b_{n-1} & c_{n-1} \\ & & & & & a_n & b_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_i \\ \vdots \\ s_{n-1} \\ s_n \end{pmatrix} \tag{10.1}$$

On pose la relation de récurrence suivante :

$$x_i = \mathrm{E}_i x_{i+1} + \mathrm{F}_i \tag{10.2}$$

En reportant l'équation (10.2) écrite pour $i-1$ dans (10.1), il vient :

$$(a_i \mathrm{E}_{i-1} + b_i) x_i + c_i x_{i+1} = s_i - a_i \mathrm{F}_{i-1} \tag{10.3}$$

en écrivant que les équations (10.2) et (10.3) sont compatibles, nous avons :

$$\frac{\mathrm{E}_i}{c_i} = -\frac{1}{a_i \mathrm{E}_{i-1} + b_i} = -\frac{\mathrm{F}_i}{s_i - a_i \mathrm{F}_{i-1}} \tag{10.4}$$

soit encore :

$$\begin{aligned} \mathrm{E}_i &= -c_i a_i \mathrm{E}_{i-1} + b_i \\ \mathrm{F}_i &= s_i - a_i \mathrm{F}_{i-1} a_i \mathrm{E}_{i-1} + b_i \end{aligned} \tag{10.5}$$

Connaissant E_1 et F_1, il est possible de calculer les coefficients E_i et F_i, pour $i = 2, n-1$ à l'aide des formules (10.5). Connaissant x_n, la formule (10.2) permet de calculer les x_i, pour $i = n-1, 1$. Les coefficients E_1 et F_1 s'obtiennent par identification entre $x_1 = \mathrm{E}_1 x_2 + \mathrm{F}_1$ et $b_1 x_1 + c_1 x_2 = s_1$, soit $\mathrm{E}_1 = -c_1/b_1$ et $\mathrm{F}_1 = s_1/b_1$, et la valeur x_n, par identification entre $x_n = \mathrm{E}_n x_{n+1} + \mathrm{F}_n$, $x_{n-1} = \mathrm{E}_{n-1} x_n + \mathrm{F}_{n-1}$ et $a_n x_{n-1} + b_n x_n = s_n$ c'est-à-dire :

$$x_n = \frac{s_n - a_n \mathrm{F}_{n-1}}{a_n \mathrm{E}_{n-1} + b_n} \tag{10.6}$$

On remarque que cela correspond à $x_n = \mathrm{F}_n$ ou encore $\mathrm{E}_n = 0$, d'où les formules générales de cet algorithme :

$$\begin{aligned}
&\mathrm{E}_1 \leftarrow -\frac{c_1}{b_1},\ \mathrm{F}_1 \leftarrow \frac{s_1}{b_1} \\
&\mathrm{E}_i \leftarrow -\frac{c_i}{a_i\mathrm{E}_{i-1}+b_i} && i=2,n-1 \\
&\mathrm{F}_i \leftarrow \frac{s_i - a_i\mathrm{F}_{i-1}}{a_i\mathrm{E}_{i-1}+b_i} && i=2,n \\
&x_n = \mathrm{F}_n \leftarrow \frac{s_n - a_n\mathrm{F}_{n-1}}{a_n\mathrm{E}_{n-1}+b_n} \\
&x_i \leftarrow \mathrm{E}_i x_{i+1} + \mathrm{F}_i && i=n-1,1
\end{aligned} \tag{10.7}$$

ou encore :

$$\begin{aligned}
&\mathrm{E}_1 \leftarrow -\frac{c_1}{b_1},\ \mathrm{F}_1 \leftarrow \frac{s_1}{b_1} \\
&\mathrm{E}_i \leftarrow -\frac{c_i}{a_i\mathrm{E}_{i-1}+b_i} && i=2,n-1 \\
&\mathrm{F}_i \leftarrow \frac{s_i - a_i\mathrm{F}_{i-1}}{a_i\mathrm{E}_{i-1}+b_i} && i=2,n \\
&\mathrm{E}_n \leftarrow 0 \\
&x_i \leftarrow \mathrm{E}_i x_{i+1} + \mathrm{F}_i && i=n,1
\end{aligned} \tag{10.8}$$

Il est alors possible de noter que :

1. en prenant 0 comme valeur des coefficients E_0 et F_0, les formules (10.5) s'appliquent aussi en $i=1$, de même, en prenant $c_n = 0$, ces formules s'appliquent pour $i = n$;
2. dans la pratique, il n'est pas utile de créer des tableaux E et F, sauf si on a à résoudre la matrice pour plusieurs seconds membres et on a :

$$\begin{aligned}
&c_1 \leftarrow -\frac{c_1}{b_1},\ d_1 \leftarrow \frac{d_1}{b_1} \\
&c_i \leftarrow -\frac{c_i}{a_i\mathrm{E}_{i-1}+b_i} && i=2,n-1 \\
&d_i \leftarrow \frac{s_i - a_i\mathrm{F}_{i-1}}{a_i\mathrm{E}_{i-1}+b_i} && i=2,n \\
&x_n \leftarrow \mathrm{F}_n = \frac{s_n - a_n\mathrm{F}_{n-1}}{a_n\mathrm{E}_{n-1}+b_n} \\
&x_i \leftarrow \mathrm{E}_i x_{i+1} + \mathrm{F}_i && i=n-1,1
\end{aligned} \tag{10.9}$$

3. L'algorithme de Thomas consiste en fait à effectuer une transformation sur le système initial comme suit :

$$\begin{bmatrix}
1 & -\mathrm{E}_1 & & & & \\
 & 1 & -\mathrm{E}_2 & & & \\
 & & \ddots & & & \\
 & & 1 & -\mathrm{E}_i & & \\
 & & & \ddots & & \\
 & & & 1 & -\mathrm{E}_{n-1} & \\
 & & & & 1 &
\end{bmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}
=
\begin{pmatrix} \mathrm{F}_1 \\ \mathrm{F}_2 \\ \vdots \\ \mathrm{F}_i \\ \vdots \\ \mathrm{F}_{n-1} \\ \mathrm{F}_n \end{pmatrix} \tag{10.10}$$

puis à inverser ce système (10.10).

10.2 Factorisation LU

Un système tridiagonal peut se résoudre de façon très rapide en utilisant la factorisation LU, en effet une matrice tridiagonale se factorise en deux matrices bidiagonales et la factorisation n'augmente donc pas le nombre de coefficients à stocker. Le système à résoudre est le système (10.1).

En appelant l_i le terme de la diagonale inférieure de la matrice L, d_i le terme de la diagonale principale et u_i le terme de la diagonale supérieure de la matrice U, la méthode s'écrit :

(a) factorisation : $d_1 = b_1$; $u_1 = c_1$

$$\begin{aligned} l_i &= a_i/d_{i-1} \\ u_i &= c_i \qquad\qquad i = 2,n \\ d_i &= b_i - l_i c_{i-1} \end{aligned} \tag{10.11}$$

(b) résolution de $\mathbf{Ly} = \mathbf{s}$: $y_1 = s_1$

$$y_i = s_i - l_i y_{i-1} \quad i = 2,n \tag{10.12}$$

(c) résolution de $\mathbf{Ux} = \mathbf{y}$: $x_n = y_n/d_n$

$$x_i = \frac{y_i - u_i x_{i+1}}{d_i} \quad i = n-1,1 \tag{10.13}$$

Le compte d'opération est le suivant :

- décomposition : $(n-1)$(1 addition +1 division +1 multiplication)
- résolution : $2(n-1)$(1 addition +1 multiplication)+ n divisions

10.3 Réduction cyclique

10.3.1 Principe

Un système tridiagonal peut aussi se résoudre par une méthode de réduction cyclique. Le principe est de réordonner la matrice initiale en fonction des puissances de 2 et d'effectuer des éliminations successives. À chaque étape, le nombre des inconnues est divisé par deux et on obtient à la fin du processus une équation à une inconnue qui permet de recalculer ensuite les inconnues qui ont été éliminées. On considère ici un système tridiagonal 6×6 afin d'expliciter les différentes étapes de la méthode. Cette méthode peut aussi s'appliquer à des systèmes quelconques :

$$\begin{bmatrix} b_1 & c_1 & 0 & 0 & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & 0 & 0 \\ 0 & a_3 & b_3 & c_3 & 0 & 0 \\ 0 & 0 & a_4 & b_4 & c_4 & 0 \\ 0 & 0 & 0 & a_5 & b_5 & c_5 \\ 0 & 0 & 0 & 0 & a_6 & b_6 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \end{pmatrix} \tag{10.14}$$

En renumérotant les termes selon la règle suivante :

1, 3, 5, 7, 9, ..., $2k+1$

2, 6, 10, 14, ..., 2^2k+2

4, 12, 20, ..., 2^3k+4

8, 24, ..., 2^4k+8

16, 48, ..., 2^5k+16

les inconnues seront donc rangées dans l'ordre $(x_1\ x_3\ x_5\ x_2\ x_6\ x_4)$ et le système initial sera alors transformé sous la forme :

$$\begin{bmatrix} b_1 & 0 & 0 & c_1 & 0 & 0 \\ 0 & b_3 & 0 & a_3 & 0 & c_3 \\ 0 & 0 & b_5 & 0 & c_5 & a_5 \\ a_2 & c_2 & 0 & b_2 & 0 & 0 \\ 0 & 0 & a_6 & 0 & b_6 & 0 \\ 0 & a_4 & c_4 & 0 & 0 & b_4 \end{bmatrix} \begin{pmatrix} x_1 \\ x_3 \\ x_5 \\ x_2 \\ x_6 \\ x_4 \end{pmatrix} = \begin{pmatrix} s_1 \\ s_3 \\ s_5 \\ s_2 \\ s_6 \\ s_4 \end{pmatrix} \tag{10.15}$$

Si on considère les trois premières lignes, les variables $(x_1\ x_3\ x_5)$ sont découplées et il est facile de les exprimer en fonctions des variables $(x_2\ x_4\ x_6)$; on peut ainsi les éliminer dans les trois dernières lignes et obtenir un système de la forme :

$$\begin{bmatrix} b_1^* & c_1^* & 0 \\ a_2^* & b_2^* & c_2^* \\ 0 & a_3^* & b_3^* \end{bmatrix} \begin{pmatrix} x_2 \\ x_4 \\ x_6 \end{pmatrix} = \begin{pmatrix} s_1^* \\ s_2^* \\ s_3^* \end{pmatrix} \tag{10.16}$$

qui est en tout point analogue au système initial avec deux fois moins de variable. Si le système était quelconque, les trois variables x_1, x_3 et x_5 seraient couplées et il faudrait donc éliminer ces variables par des méthodes classiques. En réordonnant ce système en prenant la même règle que précédemment, nous aurons :

$$\begin{bmatrix} b_1^* & 0 & c_3^* \\ 0 & b_3^* & c_1^* \\ a_2^* & c_2^* & b_2^* \end{bmatrix} \begin{pmatrix} x_2 \\ x_6 \\ x_4 \end{pmatrix} = \begin{pmatrix} s_1^* \\ s_3^* \\ s_2^* \end{pmatrix} \tag{10.17}$$

les deux premières lignes permettent d'exprimer x_2 et x_6 en fonction de x_4 et donc en reportant dans la dernière équation de déterminer la valeur de x_4. Connaissant cette valeur, on peut donc calculer celles de x_2 et x_6 puis celles de x_1, x_3 et x_5.

10.3.2 Calculs

Reprenons en détail les calculs nécessaires à l'écriture de l'algorithme de réduction cyclique :

1. élimination des variables x_1, x_3 et x_5 ; nous pouvons écrire :

$$\begin{aligned} x_1 &= \frac{s_1 - c_1 x_2}{b_1} \\ x_3 &= \frac{s_3 - a_3 x_2 - c_3 x_4}{b_3} \\ x_5 &= \frac{s_5 - a_5 x_4 - c_5 x_6}{b_5} \end{aligned} \tag{10.18}$$

En reportant dans les trois autres équations, nous aurons :

$$a_2x_1 + b_2x_2 + c_2x_3 = s_2 \tag{10.19}$$

soit :

$$\frac{a_2(s_1 - c_1x_2)}{b_1} + b_2x_2 + \frac{c_2(s_3 - a_3x_2 - c_3x_4)}{b_3} = s_2$$
$$x_2\left(b_2 - \frac{a_2c_1}{b_1} - \frac{a_3c_2}{b_3}\right) - \frac{x_4c_3c_2}{b_3} = s_2 - \frac{s_1a_2}{b_1} - \frac{s_3c_2}{b_3}$$

puis :

$$\begin{aligned} &a_4x_3 + b_4x_4 + c_4x_5 = s_4 \\ &-\frac{x_2a_2a_3}{b_1} + x_4\left(b_4 - \frac{a_4c_3}{b_3} - \frac{a_5c_5}{b_5}\right) + \frac{x_6c_5c_4}{b_5} = s_4 - \frac{s_3a_4}{b_3} - \frac{s_5c_4}{b_5} \end{aligned} \tag{10.20}$$

et enfin :

$$\begin{aligned} &a_6x_5 + b_6x_6 = s_6 \\ &-\frac{x_4a_5a_6}{b_5} + x_6\left(b_6 - \frac{c_5b_6}{b_5}\right) = s_6 - \frac{s_5a_6}{b_5} \end{aligned} \tag{10.21}$$

d'où les valeurs des coefficients a_i^*, b_i^*, c_i^* et s_i^*.

2. élimination des variables x_2 et x_6 ; nous pouvons écrire :

$$\begin{aligned} x_2 &= \frac{s_1^* - c_1^*x_4}{b_1^*} \\ x_6 &= \frac{s_3^* - a_3^*x_4}{b_3^*} \end{aligned} \tag{10.22}$$

et en reportant dans la dernière équation, nous avons :

$$a_2^*x_2 + b_2^*x_4 + c_2^*x_6 = s_2 \tag{10.23}$$

soit $b_1^{**}x_4 = s_1^{**}$ avec $b_1^{**} = b_2^* - a_2^*c_1^*/b_1^* - a_3^*c_2^*/b_3^*$ et $s_1^{**} = s_2^* - a_2^*s_1^*/b_1^* - s_3^*c_2^*/b_3^*$.

3. calcul des inconnues x_1 à x_6 en utilisant les relations précédentes :

$$\begin{aligned} x_4 &= \frac{s_1^{**}}{b_1^{**}} \\ x_2 &= \frac{s_1^* - c_1^*x_4}{b_1^*} \\ x_6 &= \frac{s_3^* - a_3^*x_4}{b_3^*} \end{aligned} \tag{10.24}$$

et :

$$\begin{aligned} x_1 &= \frac{s_1^* - c1^*x_2}{b_3^*} \\ x_3 &= \frac{s_3^* - a_3^*x_2 - c_3^*x_4}{b_3^*} \\ x_5 &= \frac{s_5^* - a_5^*x_4 - c_5^*x_6}{b_3^*} \end{aligned} \tag{10.25}$$

Le procédé ci-dessus est facilement extensible aux systèmes de dimension n.

10.4 Système pentadiagonal

Il s'agit maintenant de l'algorithme de Thomas généralisé. Considérons un système pentadiagonal, et nous allons effectuer la même démarche que dans le cas d'un système tridiagonal. Un système pentadiagonal peut s'écrire sous la forme :

$$\begin{bmatrix} b_1 & c_1 & f_1 & & & & & & \\ a_2 & b_2 & c_2 & f_2 & & & & & \\ e_3 & a_3 & b_3 & c_3 & f_3 & & & & \\ & & & \ddots & & & & & \\ & & e_i & a_i & b_i & c_i & f_i & & \\ & & & & & \ddots & & & \\ & & & & e_{n-2} & a_{n-2} & b_{n-2} & c_{n-2} & f_{n-2} \\ & & & & & e_{n-1} & a_{n-1} & b_{n-1} & c_{n-1} \\ & & & & & & e_n & a_n & b_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_i \\ \vdots \\ x_{n-1} \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_i \\ \vdots \\ s_{n-2} \\ s_{n-1} \\ s_n \end{pmatrix} \tag{10.26}$$

La première étape consiste à éliminer les coefficients e_i, ce qui revient en fait à éliminer l'inconnue correspondante x_{i-2} dans l'équation i. Le système transformé s'écrit :

$$\begin{bmatrix} b'_1 & c'_1 & f'_1 & & & & & & \\ a'_2 & b'_2 & c'_2 & f'_2 & & & & & \\ & a'_3 & b'_3 & c'_3 & f'_3 & & & & \\ & & & \ddots & & & & & \\ & & & a'_i & b'_i & c'_i & f'_i & & \\ & & & & & \ddots & & & \\ & & & & a'_{n-2} & b'_{n-2} & c'_{n-2} & f'_{n-2} & \\ & & & & & a'_{n-1} & b'_{n-1} & c'_{n-1} & \\ & & & & & & a'_n & b'_n & \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_i \\ \vdots \\ x_{n-1} \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} s'_1 \\ s'_2 \\ s'_3 \\ \vdots \\ s'_i \\ \vdots \\ s'_{n-2} \\ s'_{n-1} \\ s'_n \end{pmatrix} \tag{10.27}$$

avec :

$$\begin{aligned} & b'_1 = b_1, c'_1 = c_1 f'_1 = f_1, s'_1 = s_1 \\ & a'_2 = a_2, b'_2 = b_2, c'_2 = c_2, f'_2 = f_2, s'_2 = s_2 \\ & a'_i = a_i - \frac{e_i b'_{i-1}}{a'_{i-1}} \\ & b'_i = b_i - \frac{e_i c'_{i-1}}{a'_{i-1}} \\ & c'_i = c_i - \frac{e_i f'_{i-1}}{a'_{i-1}} \\ & f'_i = f_i \\ & s'_i = s_i - \frac{e_i s'_{i-1}}{a'_{i-1}} \end{aligned} \tag{10.28}$$

La deuxième étape consiste à éliminer le coefficient a'_i ou la variable x_{i-1}, soit :

$$\begin{bmatrix} 1 & s''_1 & f''_1 & & & & & & \\ & 1 & c''_2 & f''_2 & & & & & \\ & & 1 & c''_3 & f''_3 & & & & \\ & & & & \ddots & & & & \\ & & & & 1 & c''_i & f''_i & & \\ & & & & & & \ddots & & \\ & & & & & & 1 & c''_{n-2} & f''_{n-2} \\ & & & & & & & 1 & c''_{n-1} \\ & & & & & & & & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_i \\ \vdots \\ x_{n-1} \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} s''_1 \\ s''_2 \\ s''_3 \\ \vdots \\ s''_i \\ \vdots \\ s''_{n-2} \\ s''_{n-1} \\ s''_n \end{pmatrix} \tag{10.29}$$

avec :

$$\begin{aligned} c''_1 &= \frac{c'_1}{b'_1}, \quad f''_1 = \frac{f'_1}{b'_1}, \quad s''_1 = \frac{s'_1}{b'_1} \\ c''_i &= \frac{a'_i f''_{i-1} - c'_i}{a'_i c''_{i-1} - b'_i} \\ f''_i &= -\frac{f'_i}{a'_i c''_{i-1} - b'_i} \\ s''_i &= \frac{a'_i d''_{i-1} - s'_i}{a'_i c''_{i-1} - b'_i} \end{aligned} \tag{10.30}$$

Une fois ce calcul effectué, connaissant la valeur de x_{N}, la solution s'obtient par la formule :

$$x_i = s''_i - c''_i x_{i+1} - f''_i x_{i+2} \tag{10.31}$$

11

Équations aux dérivées partielles paraboliques - Équation d'advection-diffusion instationnaire

11.1 Problème

On se propose d'étudier une équation aux dérivées partielles du second ordre de type parabolique où les deux variables seront le temps t et une dimension d'espace x (coordonnées cartésiennes). Soit $u(x,t)$ cette fonction et l'opérateur différentiel $\mathscr{L}$ qui est défini par :

$$\mathscr{L}(u) = \frac{\partial u}{\partial t} + \mathrm{V}\frac{\partial u}{\partial x} - a\frac{\partial^2 u}{\partial x^2} \tag{11.1}$$

où on suppose que V et a sont soit des constantes, soit des fonctions du temps t et de la variable d'espace x. Dans de nombreux problèmes, V est une fonction de u ; mais dans ce cas, l'équation aux dérivées partielles devient non linéaire et doit être traitée dans un processus global itératif. La connaissance de la classification des équations aux dérivées partielles et du monde des coniques est un pré requis.

Soit à résoudre le problème suivant :

- équation locale :

$$\mathscr{L}(u) = 0 \text{ pour } t > 0 \qquad x \in]x_a \;\; x_b[\tag{11.2}$$

- condition initiale donnée :

$$u(0,x) \tag{11.3}$$

- conditions limites données :

$$\begin{cases} u(t,x_a) \\ u(t,x_b) \end{cases} \quad \text{ou} \quad \begin{cases} \dfrac{\partial u(t,x_a)}{\partial x} \\ u(t,x_b) \end{cases} \quad \text{ou} \quad \begin{cases} u(t,x_a) \\ \dfrac{\partial u(t,x_b)}{\partial x} \end{cases} \tag{11.4}$$

Le domaine de calcul, défini par $t > 0$, $x_a \leqslant x \leqslant x_b$ est discrétisé comme précédemment et on appelle Δx le pas d'espace et J le nombre de points de discrétisation en x. Les conditions limites peuvent être de type Dirichlet, Neumann ou de Robbin. Dans un premier temps, nous considérerons des conditions limites de type Dirichlet ; le traitement des conditions limites faisant apparaître les dérivées spatiales fera l'objet d'un paragraphe particulier.

On notera $\mathrm{U}_j^n \simeq u(t^n, x_j)$ la valeur discrète approchée de la fonction u au temps t^n et au point d'abscisse x_j ; l'indice n désigne donc le temps et j l'espace.

11.2 Schémas explicites

11.2.1 Définition

On considère un schéma de discrétisation faisant intervenir des points aux temps $t = t^n + \Delta t$ et $t = t^n$ et faisant intervenir les valeurs de u aux trois valeurs de x, x_{j-1}, x_j et x_{j+1}. Un schéma de discrétisation sera dit *explicite* (en trois points) si on peut exprimer ce schéma sous la forme :

$$\mathrm{U}_j^{n+1} = \mathrm{F}(\mathrm{U}_{j-1}^n, \mathrm{U}_j^n, \mathrm{U}_{j+1}^n) \tag{11.5}$$

Plus généralement ; un schéma sera dit explicite si la valeur de l'inconnue en un point ne s'exprime qu'en fonction des étapes connues à des temps antérieurs ; c'est à dire que nous aurons :

$$\mathrm{U}_j^{n+1} = \mathrm{F}(\mathrm{U}_l^k) \tag{11.6}$$

avec $k \leqslant n$ et $1 \leqslant l \leqslant \mathrm{J}$.

11.2.2 Schéma FTCS

Ce schéma (Forward Time Centered Space) est obtenu en écrivant l'opérateur différentiel au point (n, j) correspondant au temps $t = t^n$ et à l'abscisse $x = x_j$. La dérivée temporelle $\frac{\partial u}{\partial t}$, s'écrit :

$$\left(\frac{\partial u}{\partial t}\right)_j^n \simeq \frac{\mathrm{U}_j^{n+1} - \mathrm{U}_j^n}{\Delta t} \qquad \text{dérivée avant en temps} \tag{11.7}$$

alors que les dérivées spatiales s'écrivent (avec un pas constant Δx en x) :

$$\left(\frac{\partial u}{\partial x}\right)_j^n \simeq \frac{\mathrm{U}_{j+1}^n - \mathrm{U}_{j-1}^n}{2\Delta x} \qquad \text{dérivée première centrée en espace} \tag{11.8}$$

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_j^n \simeq \frac{\mathrm{U}_{j+1}^n - 2\mathrm{U}_j^n + \mathrm{U}_{j-1}^n}{\Delta x^2} \qquad \text{dérivée seconde centrée en espace} \tag{11.9}$$

En remplaçant la fonction et les dérivées par leurs approximations, il vient :

$$\frac{1}{\Delta t}\left(\mathrm{U}_j^{n+1} - \mathrm{U}_j^n\right) + \frac{\mathrm{V}}{2\Delta x}\left(\mathrm{U}_{j+1}^n - \mathrm{U}_{j-1}^n\right) = \frac{a}{\Delta x^2}\left(\mathrm{U}_{j+1}^n - 2\mathrm{U}_j^n + \mathrm{U}_{j-1}^n\right) \tag{11.10}$$

soit encore :

$$\mathrm{U}_j^{n+1} = \mathrm{U}_j^n - \frac{\mathrm{V}\Delta t}{2\Delta x}\left(\mathrm{U}_{j+1}^n - \mathrm{U}_{j-1}^n\right) + \frac{a\Delta t}{\Delta x^2}\left(\mathrm{U}_{j+1}^n - 2\mathrm{U}_j^n + \mathrm{U}_{j-1}^n\right) \tag{11.11}$$

On fait ainsi apparaître deux nombres caractéristiques sans dimension qui sont :

$$\text{nombre de Courant : } \mathrm{C} = \frac{\mathrm{V}\Delta t}{\Delta x} \qquad \text{nombre de diffusion : } \mathrm{D} = \frac{a\Delta t}{\Delta x^2}$$

Le schéma ci-dessous présente la structure de la maille de discrétisation en quatre points : Le schéma explicite ainsi défini peut se mettre sous la forme générale :

$$\mathrm{U}_j^{n+1} = a_j^+ \mathrm{U}_{j+1}^n + a_j^c \mathrm{U}_j^n + a_j^- \mathrm{U}_{j-1}^n \tag{11.12}$$

$n+1$		$\circ$	
n	$\times$	$+$	$\times$
	$j-1$	j	$j+1$

figure 11.1 - *Maille de discrétisation du schéma FTCS avec* $\circ$ *point inconnu ;* $+$ *point d'application de la discrétisation,* $\times$ *point connu*

avec $a_j^+ = \mathrm{D} - \mathrm{C}/2$, $a_j^c = 1 - 2\mathrm{D}$ et $a_j^- = \mathrm{D} + \mathrm{C}/2$. Au pas de temps $n+1$, les points sont calculés de 2 à J-1 et font intervenir les points 1 à J de l'étape connue n ; ce qui implique que les conditions limites U_1^{n+1} et U_j^{n+1} n'interviennent pas au temps $n+1$ mais au temps $n+2$. Comme le schéma est explicite, des points voisins sont indépendants et peuvent donc être calculés dans un ordre quelconque, ce qui permet de vectoriser l'algorithme de résolution sur des machines vectorielles.

11.2.3 Définitions

Consistance

Un schéma sera dit *consistant* s'il approxime l'équation aux dérivées partielles, autrement dit, que si le pas d'espace et le pas de temps tendent vers zéro, alors, le schéma discret tend vers la solution du problème continu. Pour vérifier la consistance d'un schéma numérique, on utilise les développements en séries de Taylor et on remplace chaque valeur discrète en un point par le développement de la fonction continue associée. L'analyse de la consistance conduit en fait à déterminer l'erreur de troncature du schéma ainsi que l'ordre de la précision.

Stabilité

Un schéma sera dit *stable* si les erreurs ne sont pas amplifiées. Les conditions de stabilité portent sur le schéma numérique lui-même et signifient que la solution numérique (obtenue sur une machine) doit rester proche de la solution exacte du problème discret.

Convergence

Un schéma sera dit *convergent* si la solution numérique tend vers la solution exacte du problème continu lorsque les pas de discrétisation tendent vers zéro.

Théorème de Lax

Intuitivement ; on conçoit bien que les trois propriétés précédentes soient liées entre elles et il existe un théorème dit Théorème d'équivalence de Lax qui traduit ce fait :

Pour un problème aux valeurs initiales bien posé et pour un schéma de discrétisation consistant, une condition nécessaire et suffisante pour qu'il y ait convergence est que le schéma de discrétisation soit stable.

11.2.4 Consistance du schéma FTCS

Remplaçons les valeurs discrètes de u dans l'équation générale du schéma explicite à trois points :

$$\mathrm{U}_j^{n+1} - (a_j^+ \mathrm{U}_{j+1}^n + a_j^c \mathrm{U}_j^n + a_j^- \mathrm{U}_{j-1}^n) = 0 \tag{11.13}$$

soit :

$$\begin{aligned} \mathrm{U}_j^{n+1} &\simeq u(t^{n+1},x_j) \\ &\simeq u(t^n,x_j)+\Delta t\,u_t(t^n,x_j)+\frac{\Delta t^2}{2!}u_{tt}(t^n,x_j)+\frac{\Delta t^3}{3!}u_{ttt}(t^n,x_j)+\mathcal{O}(\Delta t^4) \end{aligned} \tag{11.14}$$

$$\begin{aligned} \mathrm{U}_{j+1}^{n} &\simeq u(t^n,x_{j+1}) \\ &\simeq u(t^n,x_j)+\Delta x\,u_x(t^n,x_j)+\frac{\Delta x^2}{2!}u_{xx}(t^n,x_j)+\frac{\Delta x^3}{3!}u_{xxx}(t^n,x_j) \\ &\qquad +\frac{\Delta x^4}{4!}u_{xxxx}(t^n,x_j)+\mathcal{O}(\Delta x^5) \end{aligned} \tag{11.15}$$

$$\begin{aligned} \mathrm{U}_{j-1}^{n} &\simeq u(t^n,x_{j-1}) \\ &\simeq u(t\,n,x_j)-\Delta x\,u_x(t^n,x_j)+\frac{\Delta x^2}{2!}u_{xx}(t^n,x_j)-\frac{\Delta x^3}{3!}u_{xxx}(t_n,x_j) \\ &\qquad +\frac{\Delta x^4}{4!}u_{xxxx}(t^n,x_j)+\mathcal{O}(\Delta x^5) \end{aligned} \tag{11.16}$$

En remplaçant, nous obtenons :

$$\begin{aligned} &u(t^n,x_j)(1-a_j^+-a_j^c-a_j^-)+\Delta t\,u_t(t^n,x_j)-\Delta x\,u_x(t^n,x_j)(a_j^+-a_j^-) \\ &\quad+\frac{\Delta t^2}{2!}u_{tt}(t^n,x_j)-\frac{\Delta x^2}{2!}u_{xx}(t^n,x_j)(a_j^++a_j^-) \\ &\qquad+\frac{\Delta t^3}{3!}u_{ttt}(t^n,x_j)-\frac{\Delta x^3}{3!}u_{xxx}(t^n,x_j)(a_j^+-a_j^-) \\ &\qquad\quad+\frac{\Delta x^4}{4!}u_{xxxx}(t^n,x_j)(a_j^++a_j^-)+\mathcal{O}(\Delta x^5)+\mathcal{O}(\Delta t^4) \end{aligned} \tag{11.17}$$

soit, en particularisant au schéma FTCS :

$$a_j^++a_j^c+a_j^-=1; \qquad a_j^+-a_j^-=-\mathrm{C}; \qquad a_j^++a_j^-=2\mathrm{D} \tag{11.18}$$

D'où :

$$\begin{aligned} &\Delta t\,u_t(t^n,x_j)+\mathrm{C}\Delta x\,u_x(t^n,x_j)+\frac{\Delta t^2}{2!}u_{tt}(t^n,x_j)-\mathrm{D}\frac{\Delta x^2}{2!}u_{xx}(t^n,x_j) \\ &+\frac{\Delta t^3}{3!}u_{ttt}(t^n,x_j)+\mathrm{C}\frac{\Delta x^3}{3!}u_{xxx}(t^n,x_j)-2\mathrm{D}\frac{\Delta x^4}{4!}u_{xxxx}(t^n,x_j)=\mathcal{O}(\Delta x^5)+\mathcal{O}(\Delta t^4) \end{aligned} \tag{11.19}$$

En factorisant et en simplifiant par Δt, il vient :

$$\begin{aligned} &u_t(t^n,x_j)+\mathrm{V}u_x(t^n,x_j)-a\,u_{xx}(t^n,x_j)+\frac{\Delta t}{2!}u_{tt}(t^n,x_j)+\frac{\Delta t^2}{3!}u_{ttt}(t^n,x_j) \\ &+\mathrm{V}\frac{\Delta x^2}{3!}u_{xxx}(t^n,x_j)-a\frac{\Delta x^2}{12}u_{xxxx}(t^n,x_j)+\mathcal{O}(\Delta x^3)+\mathcal{O}(\Delta t^2)=0 \end{aligned} \tag{11.20}$$

On retrouve ainsi l'équation différentielle initiale $u_t(t^n,x_j)+\mathrm{V}u_x(t^n,x_j)-a\,u_{xx}(t^n,x_j)$ et les termes principaux de l'erreur de troncature :

$$\text{E.T.}=\frac{\Delta t}{2!}u_{tt}(t^n,x_j)+\mathrm{V}\frac{\Delta x^2}{3!}u_{xxx}(t^n,x_j)-a\frac{\Delta x^2}{12}u_{xxxx}(t^n,x_j) \tag{11.21}$$

Le schéma FTCS est donc bien consistant et son ordre est 1 en temps et 2 en espace car les termes significatifs de l'erreur de troncature sont $\mathcal{O}(\Delta t)$ et $\mathcal{O}(\Delta x^2)$ respectivement et l'erreur de troncature tend bien vers zéro lorsque Δt et Δx tendent vers zéro simultanément.

11.2.5 Méthodes à directions alternées explicites ADE

Ces méthodes (Alternative Direction Explicit) gardent le caractère explicite du schéma, mais on ne pourra pas vectoriser ces méthodes du fait des dépendances avant ou arrière. Cette méthode considère deux pas de temps, ou plutôt deux solutions différentes, calculées aux mêmes pas de temps, la solution finale étant égale à la demi somme de ces deux solutions. On considère deux solutions P_j^{n+1} et Q_j^{n+1} :

Première solution

Elle s'écrit comme suit :

$$P_j^{n+1} = P_j^n - \frac{V\Delta t}{2\Delta x}\left(P_{j+1}^n - P_{j-1}^{n+1}\right) + \frac{a\Delta t}{\Delta x^2}\left(P_{j+1}^n - P_j^{n+1} - P_j^n + P_{j-1}^{n+1}\right) \tag{11.22}$$

soit encore :

$$\left(1+\frac{a\Delta t}{\Delta x^2}\right)P_j^{n+1} = \left(\frac{V\Delta t}{2\Delta x}+\frac{a\Delta t}{\Delta x^2}\right)P_{j-1}^{n+1} + \left(1-\frac{a\Delta t}{\Delta x^2}\right)P_j^n + \left(-\frac{V\Delta t}{2\Delta x}+\frac{a\Delta t}{\Delta x^2}\right)P_{j+1}^n$$

qui est calculée pour des valeurs de j croissantes. Ainsi lorsque le point $(n+1, j)$ est calculé, la valeur $(n+1, j-1)$ est connue et est donc utilisée. En reprenant les deux nombres sans dimension C et D, l'équation (11.22) devient :

$$(1+D)P_j^{n+1} = \left(\frac{C}{2}+D\right)P_{j-1}^{n+1} + (1-D)P_j^n + \left(-\frac{C}{2}+D\right)P_{j+1}^n \tag{11.23}$$

Deuxième solution

Elle s'écrit comme suit :

$$Q_j^{n+1} = Q_j^n - \frac{V\Delta t}{2\Delta x}\left(Q_{j+1}^{n+1} - Q_{j-1}^n\right) + \frac{a\Delta t}{\Delta x^2}\left(Q_{j+1}^{n+1} - Q_j^{n+1} - Q_j^n + Q_{j-1}^n\right) \tag{11.24}$$

soit encore :

$$\left(1+\frac{a\Delta t}{\Delta x^2}\right)Q_j^{n+1} = \left(-\frac{V\Delta t}{2\Delta x}+\frac{a\Delta t}{\Delta x^2}\right)Q_{j+1}^{n+1} + \left(1-\frac{a\Delta t}{\Delta x^2}\right)Q_j^n + \left(\frac{V\Delta t}{2\Delta x}+\frac{a\Delta t}{\Delta x^2}\right)Q_{j-1}^n$$

Cette solution est calculée pour des valeurs de j décroissantes. De même, en utilisant C et D, l'équation (11.24) devient :

$$(1+D)Q_j^{n+1} = \left(-\frac{C}{2}+D\right)Q_{j+1}^{n+1} + (1-D)Q_j^n + \left(\frac{C}{2}+D\right)Q_{j-1}^n \tag{11.25}$$

Solution résultante

La solution du problème U_j^{n+1} est définie par :

$$U_j^{n+1} = \frac{P_j^{n+1} + Q_j^{n+1}}{2} \tag{11.26}$$

Il est donc important de choisir des conditions initiales pour P_j^{n+1} et Q_j^{n+1} qui satisfont cette dernière relation. On peut montrer, dans le cas de l'équation de la chaleur que la précision de ce schéma est bien en $\mathcal{O}(\Delta x^2)$, mais est aussi en $\mathcal{O}(\Delta t^2)$, alors que chacune des solution est précise en $\mathcal{O}(\Delta x^2)$, $\mathcal{O}(\Delta t^2)$ et $\mathcal{O}(\frac{\Delta t}{\Delta x^2})$. En effet ; si on calcule l'erreur de troncature [1] au point $(n+1/2, j)$, nous avons pour la première solution :

$$S = \Delta t\left[\frac{\partial u}{\partial t} + V\frac{\partial u}{\partial x} - a\frac{\partial^2 u}{\partial x^2} + \frac{\Delta t}{\Delta x}\left(-\frac{V}{2}\frac{\partial u}{\partial t} + a\frac{\partial^2 u}{\partial t\,\partial x}\right) + \mathcal{O}(\Delta x^2) + \mathcal{O}(\Delta t^2)\ldots\right] \tag{11.27}$$

1. La démonstration générale du calcul de l'erreur de troncature sera effectuée dans le cas des schémas implicites.

où l'on constate la présence de termes en $\frac{\Delta t}{\Delta x}$ qui implique donc la non consistance de la première solution. Pour la deuxième solution, nous aurons :

$$\mathrm{T} = \Delta t \left[\frac{\partial u}{\partial t} + \mathrm{V}\frac{\partial u}{\partial x} - a\frac{\partial^2 u}{\partial x^2} + \frac{\Delta t}{\Delta x}\left(\frac{\mathrm{V}}{2}\frac{\partial u}{\partial t} - a\frac{\partial^2 u}{\partial t \partial x} \right) + \mathcal{O}(\Delta x^2) + \mathcal{O}(\Delta t^2)\ldots \right] \tag{11.28}$$

Dans cette expression ; il y a aussi des termes en $\frac{\Delta t}{\Delta x}$ et la encore il n'y a pas consistance. En effectuant la demi somme S+T ; nous obtenons :

$$\frac{\mathrm{S}+\mathrm{T}}{2} = \Delta t \left[\frac{\partial u}{\partial t} + \mathrm{V}\frac{\partial u}{\partial x} - a\frac{\partial^2 u}{\partial x^2} + \mathcal{O}(\Delta x^2) + \mathcal{O}(\Delta t^2)\ldots \right] \tag{11.29}$$

et on voit que les termes en $\frac{\Delta t}{\Delta x}$ ont été supprimés. Le schéma ADE est donc consistant. Il est intéressant de noter qu'ici, on a obtenu un schéma consistant à partir de deux étapes qui sont individuellement non consistantes.

Forme générale

Les deux schémas précédents peuvent se mettre sous la forme générale suivante :

$$\alpha_j^+ \mathrm{U}_{j+1}^{n+1} + \alpha_j^c \mathrm{U}_j^{n+1} + \alpha_j^- \mathrm{U}_{j-1}^{n+1} = a_j^+ \mathrm{U}_{j+1}^n + a_j^c \mathrm{U}_j^n + a_j^- \mathrm{U}_{j-1}^n \tag{11.30}$$

Ceci entraîne, pour la solution P :

$$(1+\mathrm{D})\mathrm{P}_j^{n+1} - \left(\frac{\mathrm{C}}{2} + \mathrm{D}\right)\mathrm{P}_{j-1}^{n+1} = (1-\mathrm{D})\mathrm{P}_j^n + \left(-\frac{\mathrm{C}}{2} + \mathrm{D}\right)\mathrm{P}_{j+1}^n \tag{11.31}$$

avec $\alpha_j^+ = 0$, $\alpha_j^c = 1+\mathrm{D}$, $\alpha_j^- = -\mathrm{D} - \mathrm{C}/2$, $a_j^+ = \mathrm{D} - \mathrm{C}/2$, $a_j^c = 1-\mathrm{D}$ et $a_j^- = 0$. Concernant la solution Q, on trouve :

$$\left(\frac{\mathrm{C}}{2} - \mathrm{D}\right)\mathrm{Q}_{j+1}^{n+1} + (1+\mathrm{D})\mathrm{Q}_j^{n+1} = (1-\mathrm{D})\mathrm{Q}_j^n + \left(\frac{\mathrm{C}}{2} + \mathrm{D}\right)\mathrm{Q}_{j-1}^n \tag{11.32}$$

avec $\alpha_j^+ = \mathrm{C}/2 - \mathrm{D}$, $\alpha_j^c = 1+\mathrm{D}$, $\alpha_j^- = 0$, $a_j^+ = 0$, $a_j^c = 1-\mathrm{D}$ et $a_j^- = \mathrm{D} + \mathrm{C}/2$.

11.3 Stabilité d'un schéma explicite

11.3.1 Méthode de Von Neumann

Principe

Pour étudier la stabilité d'un schéma numérique, Von Neumann a développé une méthode basée sur l'étude de l'amplification des erreurs par un schéma numérique. On suppose que U_j^n est la solution numérique du problème discret et que $\hat{\mathrm{U}}_j^n$ est la solution exacte de ce problème discret. On ne considère ici que les solutions du problème discret et on ne compare pas ces solutions avec la solution exacte du problème continu $u(t^n, x_j)$. On peut donc définir une équation introduisant l'erreur, on pose $\mathrm{U}_j^n = \hat{\mathrm{U}}_j^n + e_j^n$ en tout point et on reporte ces valeurs dans l'équation aux différences $\mathrm{U}_j^{n+1} = a_j^+ \mathrm{U}_{j+1}^n + a_j^c \mathrm{U}_j^n + a_j^- \mathrm{U}_{j-1}^n$, soit :

$$\hat{\mathrm{U}}_j^{n+1} + e_j^{n+1} = a_j^+ \hat{\mathrm{U}}_{j+1}^n + a_j^c \hat{\mathrm{U}}_j^n + a_j^- \hat{\mathrm{U}}_{j-1}^n + a_j^+ e_{j+1}^n + a_j^c e_j^n + a_j^- e_{j-1}^n \tag{11.33}$$

Par définition de $\hat{\mathrm{U}}$, qui est solution exacte du problème discret ; nous avons une équation reliant les erreurs, à savoir :

$$e_j^{n+1} = a_j^+ e_{j+1}^n + a_j^c e_j^n + a_j^- e_{j-1}^n \tag{11.34}$$

et on peut constater que les évolutions de cette erreur sont analogues à celles de la solution. Si l'opérateur de discrétisation est linéaire, on peut écrire les différents résultats sous forme vectorielle, en posant :

- $\hat{\mathbf{U}}^n$, vecteur solution exacte au temps n ;
- $\mathbf{U}^n$, vecteur solution numérique au temps n ;
- $\mathbf{e}^n$, vecteur erreur entre solution exacte et solution numérique au temps n.

Par définition, nous aurons $\hat{\mathbf{U}}^n = \mathbf{U}^n + \mathbf{e}^n$ et on suppose que l'équation aux différences se met sous forme matricielle $\mathbf{U}^{n+1} = \mathbf{A}\mathbf{U}^n$ où $\mathbf{A}$ est une matrice carrée. Nous aurons donc :

$$\hat{\mathbf{U}}^{n+1} - \mathbf{e}^{n+1} = \mathbf{A}\hat{\mathbf{U}}^n - \mathbf{A}\mathbf{e}^n \tag{11.35}$$

et comme $\hat{\mathbf{U}}^{n+1} = \mathbf{A}\hat{\mathbf{U}}^n$, nous avons aussi $\mathbf{e}^{n+1} = \mathbf{A}\mathbf{e}^n$. L'erreur et la solution dépendent donc du même opérateur linéaire $\mathbf{A}$.

Décomposition de l'erreur en séries de Fourier

La méthode de Von Neumann va supposer qu'il est possible de décomposer dans l'espace en séries de Fourier l'erreur en introduisant un nombre d'onde $k = 2\pi/\lambda$ où λ est une longueur d'onde caractéristique. Le signal sur l'intervalle $[x_a \quad x_b]$ est ramené sur l'intervalle $[0 \ \mathrm{L}]$ où $\mathrm{L} = x_b - x_a$ et prolongé par symétrie sur l'intervalle $[-\mathrm{L} \ \mathrm{L}]$. Si Δx est le pas de discrétisation ; la plus petite longueur d'onde admissible est $\lambda_{\min} = 2\Delta x$ et la plus grande longueur d'onde sera $\lambda_{\max} = 2\mathrm{L}$. Si maintenant on raisonne sur les nombres d'onde, nous aurons $k_{\max} = \pi/\Delta x$ et $k_{\min} = \pi/\mathrm{L}$. Dans la décomposition de Fourier, si j désigne l'indice du point les harmoniques seront donnés par la formule pour $j = 0, \mathrm{J}-1$:

$$k_j = j\,k_{\min} = \frac{j\,\pi}{\mathrm{L}} = \frac{j\,\pi}{(\mathrm{J}-1)\Delta x} \qquad j = 0, \mathrm{J}-1 \tag{11.36}$$

On a supposé que le maillage contenait J points en x et que $\Delta x = \mathrm{L}/(\mathrm{J}-1)$.

Dans ces conditions, nous aurons comme expression de l'erreur :

$$e_j^n = \sum_{l=-\mathrm{J}+1}^{\mathrm{J}-1} \mathrm{E}_l^n \mathrm{e}^{i k_l j \Delta x} = \sum_{l=-\mathrm{J}+1}^{\mathrm{J}-1} \mathrm{E}_l^n \mathrm{e}^{i j l \pi/(\mathrm{J}-1)} \tag{11.37}$$

avec $\mathrm{i}^2 = -1$ et E_l^n désigne l'amplitude au temps n de la composante l de la série. L'harmonique correspondant à $l = 0$ est une fonction constante dans l'espace et correspond à la valeur moyenne du signal. Le terme $k_l \Delta x$ est appelé angle de phase et noté : $\theta = k_l \Delta x = l\,\pi(\mathrm{J}-1)$; pour θ voisin de zéro, les fréquences concernées seront faibles, alors que pour θ voisin de π, les fréquences seront élevées.

En reportant cette expression dans l'équation aux différences régissant l'erreur, $e_j^{n+1} = a_j^+ e_{j+1}^n + a_j^c e_j^n + a_j^- e_{j-1}^n$, nous avons :

$$\sum_{l=-\mathrm{J}+1}^{\mathrm{J}-1} \mathrm{E}_l^{n+1} \mathrm{e}^{i k_l j \Delta x} = \sum_{l=-\mathrm{J}+1}^{\mathrm{J}-1} \mathrm{E}_l^n \left(a_j^+ \mathrm{e}^{i k_l (j+1) \Delta x} + a_j^c \mathrm{e}^{i k_l j \Delta x} + a_j^- \mathrm{e}^{i k_l (j-1) \Delta x} \right) \tag{11.38}$$

soit en divisant par $\mathrm{e}^{i k_l j \Delta x}$:

$$\sum_{l=-\mathrm{J}+1}^{\mathrm{J}-1} \mathrm{E}_l^{n+1} = \sum_{l=-\mathrm{J}+1}^{\mathrm{J}-1} \mathrm{E}_l^n \left(a_j^+ \mathrm{e}^{i\theta} + a_j^c + a_j^- \mathrm{e}^{-i\theta} \right) \tag{11.39}$$

Module d'amplification

En isolant une composante l :

$$\mathrm{E}_l^{n+1} = a_j^+ \mathrm{E}_l^n e^{\mathrm{i}\theta} + a_j^c \mathrm{E}_l^n + a_j^- \mathrm{E}_l^n e^{-\mathrm{i}\theta} \tag{11.40}$$

que l'on peut mettre sous la forme $\mathrm{E}^{n+1} = g\mathrm{E}^n$, où g est le facteur d'amplification de l'erreur en fonction du temps pour la composant l de la transformée de Fourier. En écrivant que le module de g est inférieur ou égal à 1 pour toutes les composantes et quelle que soit la valeur de l on dira que le schéma est stable.

$$|g| = \left| \frac{\mathrm{E}^{n+1}}{\mathrm{E}^n} \right| \leqslant 1 \tag{11.41}$$

Calculons le module d'amplification g :

$$\mathrm{E}_l^{n+1} = a_j^+ \mathrm{E}_l^n e^{\mathrm{i}\theta} + a_j^c \mathrm{E}_l^n + a_j^- \mathrm{E}_l^n e^{-\mathrm{i}\theta} \tag{11.42}$$

soit en omettant l'indice l :

$$\mathrm{E}^{n+1} = \mathrm{E}^n (a_j^+ e^{\mathrm{i}\theta} + a_j^c + a_j^- e^{-\mathrm{i}\theta}) \tag{11.43}$$

d'où la valeur de g :

$$g = a_j^c + (a_j^+ + a_j^-)\cos\theta + \mathrm{i}(a_j^+ - a_j^-)\sin\theta \tag{11.44}$$

g est un nombre complexe dont il faut étudier le module. Ce nombre complexe dépend des deux pas de discrétisation temporelle Δt et spatiale Δx qui sont choisis par l'utilisateur du schéma numérique et des constantes physiques du problème (V et a dans l'exemple).

Étude graphique du module d'amplification

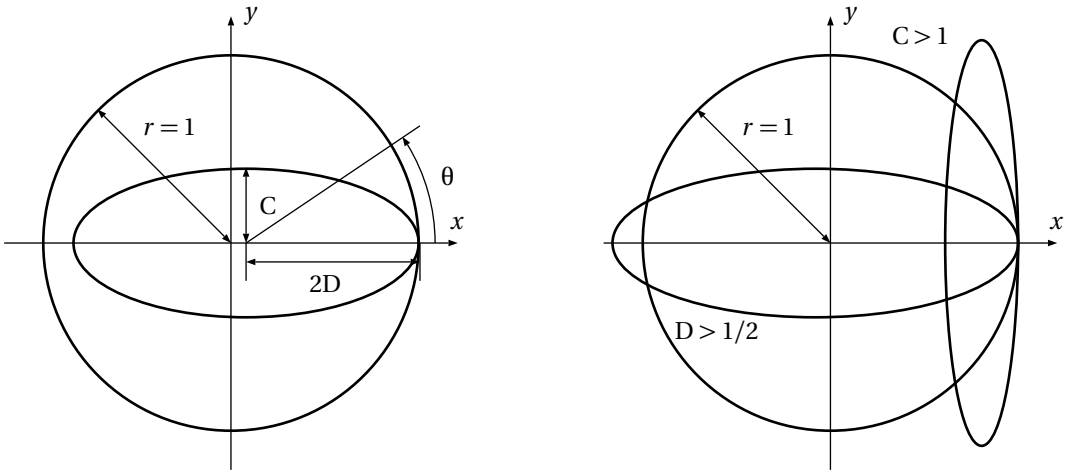
La manière la plus simple, lorsque les valeurs des coefficients a est compliquée, est d'étudier graphiquement le problème et d'étudier la position du point d'affixe g dans le plan complexe pour toutes les valeurs de θ telles que $0 \leqslant \theta \leqslant 2\pi$. Si l'ensemble des points P est situé à l'intérieur du cercle unité, alors le schéma sera stable. On peut remarquer que g peut être représenté par une ellipse définie par la courbe paramétrique (en fonction de θ) :

$$\begin{aligned} x &= a_j^c + (a_j^+ + a_j^-)\cos\theta = 1 - (a_j^+ + a_j^-)(1 - \cos\theta) \\ y &= (a_j^+ - a_j^-)\sin\theta \end{aligned} \tag{11.45}$$

Cette ellipse est centrée au point $(a_j^c, 0)$, et a pour axes $|a_j^+ + a_j^-|$ et $|a_j^+ - a_j^-|$. Dans le cas du schéma FTCS, nous avons :

$$\begin{aligned} a_j^c &= 1 - (a_j^+ + a_j^-) = 1 - 2\mathrm{D} \\ a_j^+ + a_j^- &= 2\mathrm{D} \\ a_j^+ - a_j^- &= -\mathrm{C} \end{aligned} \tag{11.46}$$

Par conséquent, l'ellipse est centrée au point $(1 - 2\mathrm{D}, 0)$ et a pour axes 2D et C. Les figures 11.2(a) et 11.2(b)expliquent ces phénomènes pour le schéma FTCS.



(a) $1-2D \geqslant 0$ et $|C| \leqslant 1$ (b) $1-2D \leqslant 0$ et $|C| \geqslant 1$

figure 11.2 - *Module d'amplification pour le schéma FTCS*

Géométriquement, les deux conditions nécessaires suivantes apparaissent :

$$1-2D \geqslant 0 \qquad \text{et} \qquad |C| \leqslant 1 \tag{11.47}$$

Ces conditions *ne sont pas suffisantes*, en effet, pour être complet, il faut aussi imposer que l'ellipse est inclue dans le cercle de rayon unité. Il faut donc calculer les intersections de l'ellipse et du cercle et écrire que la seule solution est le point $(1,0)$.

L'équation de cette ellipse est :

$$\frac{(x-(1-a_j^+-a_j^-))^2}{(a_j^++a_j^-)^2}+\frac{y^2}{(a_j^+-a_j^-)^2}=1 \tag{11.48}$$

et l'équation du cercle est :

$$x^2+y^2=1 \tag{11.49}$$

En éliminant y^2 entre les deux équations, on trouve :

$$x=\frac{(1-2(a_j^++a_j^-))(a_j^+-a_j^-)^2+(a_j^++a_j^-)^2}{(a_j^+-a_j^-)^2-(a_j^++a_j^-)^2} \tag{11.50}$$

et cette valeur doit être telle que $|x|<1$ pour qu'il y ait stabilité. On trouve alors la condition :

$$(a_j^+-a_j^-)^2 \leqslant a_j^++a_j^- \tag{11.51}$$

soit, dans le cas précis du schéma FTCS :

$$C^2 \leqslant 2D \tag{11.52}$$

qui vient s'ajouter à la condition $D \leqslant 1/2$. Finalement, on obtient les conditions suivant :

$$C^2 \leqslant 2D \leqslant 1 \tag{11.53}$$

En recombinant ces conditions, la condition nécessaire $|C| \leqslant 1$ est bien retrouvée.

Étude analytique du module d'amplification

On peut écrire $g^2 = (a_j^c + (a_j^+ + a_j^-)\cos\theta)^2 + ((a_j^+ - a_j^-)\sin\theta)^2$, et obtenir ainsi un polynôme de degré 2 en $\cos\theta$ et calculer les variations de cette fonction. Le calcul est souvent fastidieux et il est préférable, sauf cas particulier d'utiliser une autre méthode d'analyse.

Étude analytique du module d'amplification avec transformation

Écrivons :

$$g^2 = (a_j^c + (a_j^+ + a_j^-)\cos\theta)^2 + ((a_j^+ - a_j^-)\sin\theta)^2 \tag{11.54}$$

et développons cette expression ; il vient :

$$g^2 = (a_j^c)^2 + (a_j^+ + a_j^-)^2\cos^2\theta + 2a_j^c(a_j^+ + a_j^-)\cos\theta + (a_j^+ - a_j^-)^2\sin^2\theta \tag{11.55}$$

Effectuons la transformation :

$$z = \frac{1-\cos\theta}{2} \tag{11.56}$$

où z varie de façon monotone entre 0 et 1 lorsque θ varie sur un intervalle de 2π. Nous avons donc $\cos\theta = 1-2z$, $\cos^2\theta = (2z-1)^2$ et $\sin^2\theta = 4z(1-z)$; en remplaçant dans l'expression de g^2, il vient :

$$\begin{aligned} g^2 = {} & (a_j^c)^2 + (a_j^+ + a_j^-)^2 + 2a_j^c(a_j^+ + a_j^-) + 4z - (a_j^+ + a_j^-)^2 + (a_j^+ - a_j^-)^2 \\ & - a_j^c(a_j^+ + a_j^-) + 4z^2(a_j^+ + a_j^-)^2 + (a_j^+ - a_j^-)^2 \end{aligned} \tag{11.57}$$

soit encore :

$$g^2 = (a_j^+ + a_j^c + a_j^-)^2 - 4z4a_j^+a_j^- + a_j^c(a_j^+ + a_j^-) + 16z^2a_j^+a_j^- \tag{11.58}$$

Lorsque l'opérateur différentiel ne comporte pas de terme « source » en u ; on vérifie que $a_j^+ + a_j^c + a_j^- = 1$ et dans ce cas, le module de g se simplifie :

$$g^2 = 1 + 4z(a_j^+ - a_j^-)^2 - (a_j^+ + a_j^-) + 16z^2a_j^+a_j^- \tag{11.59}$$

et écrire que $|g| \leqslant 1$, équivaut à imposer :

$$4z(a_j^+ - a_j^-)^2 - (a_j^+ + a_j^-) + 16z^2a_j^+a_j^- \leqslant 0 \tag{11.60}$$

or comme z est compris entre 0 et 1, pour satisfaire la condition ; il suffit d'écrire que :

$$(a_j^+ - a_j^-)^2 - (a_j^+ - a_j^-) + 4za_j^+a_j^- \leqslant 0 \tag{11.61}$$

d'où les conditions $(a_j^+ - a_j^-)^2 - (a_j^+ + a_j^-) \leqslant 0$ en $z = 0$ et $(a_j^+ - a_j^-)^2 - (a_j^+ - a_j^-) + 4a_j^+a_j^- \leqslant 0$ en $z = 1$. Finalement :

$$\begin{aligned} & (a_j^+ - a_j^-)^2 - (a_j^+ + a_j^-) \leqslant 0 \\ & (a_j^+ + a_j^-)^2 - (a_j^+ + a_j^-) \leqslant 0 \end{aligned} \tag{11.62}$$

ou encore :

$$\begin{aligned} & (a_j^+ - a_j^-)^2 - (a_j^+ + a_j^-) \leqslant 0 \\ & (a_j^+ + a_j^-)(a_j^+ + a_j^- - 1) = -a_j^c(a_j^+ + a_j^-) \leqslant 0 \end{aligned} \tag{11.63}$$

11.3.2 Application au schéma FTCS

Nous allons utiliser les conditions précédentes pour analyser la stabilité du schéma FTCS. Les coefficients sont $a_j^+ = \mathrm{D} - \mathrm{C}/2$, $a_j^c = 1 - 2\mathrm{D}$ et $a_j^- = \mathrm{D} + \mathrm{C}/2$ et nous vérifions donc que $a_j^+ + a_j^c + a_j^- = 1$; dans les conditions de stabilité les coefficients interviennent par la somme et la différence de a_j^+ et a_j^-, soit $a_j^+ + a_j^- = 2\mathrm{D}$ et $a_j^+ - a_j^- = -\mathrm{C}$; en remplaçant dans les deux conditions précédentes, nous avons :

$$\mathrm{C}^2 - 2\mathrm{D} \leqslant 0 \qquad \text{et} \qquad 4\mathrm{D}^2 - 2\mathrm{D} \leqslant 0 \tag{11.64}$$

ce qui fournit comme conditions sur C et D :

$$\mathrm{D} \leqslant 1/2 \qquad \text{et} \qquad \mathrm{C}^2 \leqslant 2\mathrm{D} \tag{11.65}$$

En combinant ces deux conditions, nous avons alors :

$$\mathrm{D} \leqslant 1/2 \qquad \text{et} \qquad \mathrm{C}^2 \leqslant 2\mathrm{D} \leqslant 1 \tag{11.66}$$

La condition $\mathrm{C}^2 \leqslant 1$ est la condition classique de Courant Friedrich Levy dite condition de CFL. Cette condition ne porte que sur le terme de convection car il fait intervenir la vitesse.

Examinons ce qui se passe pour Δt et Δx :

(i) $\mathrm{D} \leqslant 1/2$, $\frac{a\Delta t}{\Delta x^2} \leqslant 1/2$ s'où $\Delta t \leqslant \frac{\Delta x^2}{2a}$: les pas d'espace et de temps sont liés ;

(ii) $\mathrm{C}^2 \leqslant 2\mathrm{D}$, $\Delta t \leqslant \frac{a}{\mathrm{V}^2}$: le pas d'espace n'intervient pas ;

(iii) $\mathrm{C}^2 \leqslant 1$, $\Delta t \leqslant \frac{\Delta x}{|\mathrm{V}|}$.

Ces trois conditions sont de nature différente ; en effet, la première condition ne fait intervenir que le pas de temps et le coefficient de diffusion ; la seconde condition ne fait pas intervenir le pas d'espace. Si on calcule le rapport $\frac{|\mathrm{C}|}{\mathrm{D}}$, on peut définir un nombre sans dimension appelé par analogie nombre de Reynolds de maille, soit :

$$\mathrm{R_{ex}} = \frac{|\mathrm{C}|}{\mathrm{D}} = \frac{|\mathrm{V}|\Delta x}{a} \tag{11.67}$$

et on a alors comme condition :

$$\frac{\mathrm{R_{ex}}^2}{\mathrm{D}^2} \leqslant 2\mathrm{D} \tag{11.68}$$

ou encore :

$$\mathrm{D} \leqslant \frac{2}{\mathrm{R_{ex}}^2} \quad \text{et} \quad \mathrm{D} \leqslant \frac{1}{2} \tag{11.69}$$

Si $\mathrm{R_{ex}}^2 \leqslant 2$, la condition la plus contraignante est $\mathrm{D} \leqslant 1/2$; par contre si $\mathrm{R_{ex}}^2 > 2$, c'est cette condition qui est la plus stricte.

11.3.3 Méthode d'analyse matricielle

Nous allons utiliser une autre méthode pour analyser la stabilité du schéma FTCS, cette méthode est basée sur le calcul des valeurs propres d'une matrice. L'erreur suivant la même évolution que la solution ; on peut écrire :

$$e_j^{n+1} = a_j^+ e_{j+1}^n + a_j^c e_j^n + a_j^- e_{j-1}^n \qquad j = 2, \mathrm{J}-1 \tag{11.70}$$

si on suppose qu'aux deux extrémités, l'erreur est nulle, nous aurons :

$$\begin{aligned} e_2^{n+1} &= a_2^c e_2^n + a_2^+ e_3^n \\ e_3^{n+1} &= a_3^- e_2^n + a_3^c e_3^n + a_3^+ e_4^n \\ &\vdots \\ e_j^{n+1} &= a_j^- e_{j-1}^n + a_j^c e_j^n + a_j^+ e_{j+1}^n \\ e_{j-1}^{n+1} &= a_{j-1}^- e_{j-2}^n + a_{j-1}^c e_{j-1}^n \end{aligned} \tag{11.71}$$

Et on peut trouver une forme matricielle $\mathbf{e}^{n+1} = \mathbf{A}\mathbf{e}^n$, où $\mathbf{A}$ est une matrice carrée d'ordre J−2. L'erreur $\mathbf{e}^{n+1}$ est bornée si le rayon spectral de la matrice $\mathbf{A}$ est inférieur à 1. Dans le cas d'une équation de diffusion pure, la matrice associée a la forme suivante :

$$\mathbf{A} = \begin{bmatrix} 1-2\mathrm{D} & \mathrm{D} & & & & & & \\ \mathrm{D} & 1-2\mathrm{D} & \mathrm{D} & & & & & \\ & & \ddots & & & & & \\ & & \mathrm{D} & 1-2\mathrm{D} & \mathrm{D} & & & \\ & & & & \ddots & & & \\ & & & & \mathrm{D} & 1-2\mathrm{D} & \mathrm{D} & \\ & & & & & \mathrm{D} & 1-2\mathrm{D} \end{bmatrix} \tag{11.72}$$

Les J−2 valeurs propres de cette matrice sont calculables analytiquement et nous avons :

$$\lambda_j = 1 - 4\mathrm{D}\sin^2\frac{j\pi}{2(\mathrm{J}-3)} \qquad j = 0, \mathrm{J}-3 \tag{11.73}$$

Le rayon spectral de la matrice $\mathbf{A}$, noté $\rho(\mathbf{A})$ est le maximum sur j de la valeur absolue de λ_j, soit :

$$\rho(\mathbf{A}) = \max|\lambda_j| \tag{11.74}$$

en effet ; les deux valeurs extrêmes des valeurs propres sont respectivement 1 et $1-4\mathrm{D}$. La condition $-1 \leqslant 1-4\mathrm{D} \leqslant 1$, se traduit par $\mathrm{D} \leqslant 1/2$. Pour que $\rho(\mathbf{A})$ soit inférieur ou égal à 1 quelque soit j, il faut donc que D soit inférieur à 1/2.

Dans le cas où les coefficients a_j^+, a_j^- et a_j^c ne sont pas constants, il faut calculer numériquement la valeur propre de plus grand module et vérifier que son module est inférieur à 1.

11.4 Schémas explicites décentrés

11.4.1 Décentration

Dans le chapitre précédent ; nous avons vu que le fait de décentrer la dérivée première en fonction du signe de V permettait d'éviter les oscillations spatiales (admissibilité des racines). Il est donc souhaitable d'étudier l'influence de la décentration sur les conditions générale de stabilité. On rappelle que nous utiliserons une décentration « upwind » qui dépend de la valeur de V ; ainsi :

– $\mathrm{V} > 0$:

$$\left(\frac{\partial u}{\partial x}\right)_j^n \simeq \frac{\mathrm{U}_j^n - \mathrm{U}_{j-1}^n}{\Delta x} \tag{11.75}$$

– $V<0$:

$$\left(\frac{\partial u}{\partial x}\right)_j^n \simeq \frac{U_{j+1}^n - U_j^n}{\Delta x} \tag{11.76}$$

d'où la forme générale :

$$\left(\frac{\partial u}{\partial x}\right)_j^n \simeq \frac{\alpha U_{j+1}^n + (1-2\alpha)U_j^n - (1-\alpha)U_{j-1}^n}{\Delta x} \tag{11.77}$$

avec $\alpha = 0$ si $V > 0$ ($C > 0$) et $\alpha = 1$ si $V < 0$ ($C < 0$). Dans ces conditions, le schéma décentré s'écrit :

$$U_j^{n+1} = a_j^+ U_{j+1}^n + a_j^c U_j^n + a_j^- U_{j-1}^n \tag{11.78}$$

avec :

$$\begin{aligned}
a_j^+ &= -\alpha\frac{V\Delta t}{\Delta x} + \frac{a\Delta t}{\Delta x^2} = D - \alpha C \\
a_j^c &= 1 - (1-2\alpha)\frac{V\Delta t}{\Delta x} - 2\frac{a\Delta t}{\Delta x^2} = 1 - (1-2\alpha)C - 2D \\
a_j^- &= (1-\alpha)\frac{V\Delta t}{\Delta x} + \frac{a\Delta t}{\Delta x^2} = D + (1-\alpha)C
\end{aligned}$$

En utilisant les résultats précédents concernant la stabilité, on montre que $a_j^+ + a_j^- = 2D + (1-2\alpha)C$ et $a_j^+ - a_j^- = -C$, d'où les deux conditions :

$$C^2 - 2D - (1-2\alpha)C \leqslant 0 \qquad \text{et} \qquad (2D + (1-2\alpha C))(1 - 2D - (1-2\alpha C)) \leqslant 0 \tag{11.79}$$

Examinons la première condition :

– si $\alpha = 0$, nous avons $C^2 - C < 2D$;
– si $\alpha = 1$, nous avons $C^2 + C < 2D$.

Ces deux cas peuvent se mettre sous la même forme, soit :

$$C^2 - |C| < 2D \tag{11.80}$$

La deuxième condition s'écrira :

– si $\alpha = 0$, $(2D + C)(1 - 2D - C) \leqslant 0$;
– si $\alpha = 1$, $(2D - C)(1 - 2D + C) \leqslant 0$.

soit la forme générale :

$$(2D + |C|)(1 - 2D - |C|) \leqslant 0 \tag{11.81}$$

comme le premier terme est nécessairement positif, la condition s'écrit :

$$1 - 2D - |C| \leqslant 0 \tag{11.82}$$

Ces conditions sont plus difficiles à interpréter directement ; aussi il est pratique d'introduire le nombre de Reynolds de maille R_{ex}, soit :

$$D \leqslant \frac{2 + R_{ex}}{{R_{ex}}^2}; \qquad D \leqslant \frac{1}{2 + R_{ex}} \tag{11.83}$$

La première constatation (paradoxale) est que, pour de faibles valeurs du nombre de Reynolds de maille, la décentration « upwind » a un effet pénalisant sur les conditions de stabilité, par contre pour les fortes valeurs, le fait de décentrer impose des conditions de restriction moindres. Si on compare les deux schémas (centré et décentré), la valeur critique du nombre de Reynolds de maille pour laquelle le schéma décentré est moins pénalisant s'obtient par l'équation :

$$\frac{1}{2+\mathrm{R_{ex}}}=\frac{1}{{\mathrm{R_{ex}}}^2} \quad \Rightarrow \quad \mathrm{R_{ex}}=1+\sqrt{5} \tag{11.84}$$

11.5 Schémas implicites

11.5.1 Définition

On considère un schéma de discrétisation faisant intervenir des points aux temps $t = t^{n+1} = t^n + \Delta t$ et $t = t^n$ et aux valeurs de x, x_{j-1}, x_j et x_{j+1}. Un schéma de discrétisation sera dit *implicite* si on peut exprimer ce schéma sous la forme :

$$\mathrm{G}(\mathrm{U}_{j-1}^{n+1},\mathrm{U}_j^{n+1},\mathrm{U}_{j+1}^{n+1})=\mathrm{F}(\mathrm{U}_{j-1}^{n},\mathrm{U}_j^{n},\mathrm{U}_{j+1}^{n}) \tag{11.85}$$

Un schéma sera dit implicite si la valeur de l'inconnue en un point s'exprime en fonction de celles de points voisins à cette étape inconnue et à des temps antérieurs. Il n'est donc pas possible de calculer un point $(n+1, j)$ de façon isolée. Le fait d'introduire des schémas implicites doit permettre d'améliorer les propriétés des schémas explicites et ce de deux façons, soit en augmentant la précision, soit en augmentant le domaine de stabilité.

11.5.2 Schéma de Crank-Nicolson

Ce schéma est obtenu en écrivant l'opérateur différentiel en un point correspondant à un temps intermédiaire $t^* = t^n + \Delta t/2$ et à l'abscisse $x = x_j$. La dérivée temporelle $\frac{\partial u}{\partial t}$, s'écrit :

$$\left(\frac{\partial u}{\partial t}\right)_j^{n+1/2} \simeq \frac{\mathrm{U}_j^{n+1}-\mathrm{U}_j^n}{\Delta t} \quad \text{dérivée centrée en temps} \tag{11.86}$$

et les dérivées spatiales seront des pondérations entre les temps n et $n+1$; soit :

$$\left(\frac{\partial u}{\partial x}\right)_j^{n+1/2}=\theta\left(\frac{\partial u}{\partial x}\right)_j^{n+1}+(1-\theta)\left(\frac{\partial u}{\partial x}\right)_j^{n} \qquad 0\leqslant\theta\leqslant 1 \tag{11.87}$$

Plus généralement ; on peut écrire l'équation aux dérivées partielles initiale sous la forme :

$$\frac{\partial u}{\partial t}=\mathscr{L}_x(u) \tag{11.88}$$

où $\mathscr{L}_x$ est l'opérateur différentiel spatial défini par :

$$\mathscr{L}_x(u)=-\mathrm{V}\frac{\partial u}{\partial x}+a\frac{\partial^2 u}{\partial x^2} \tag{11.89}$$

Dans ce cas, on écrira

$$\left(\frac{\partial u}{\partial t}\right)_j^{n+1/2}=\mathscr{L}_x(u)|_j^{n+1/2}=\theta\mathscr{L}_x(u)|_j^{n+1}+(1-\theta)\mathscr{L}_x(u)|_j^{n} \tag{11.90}$$

Le schéma de Crank Nicolson proprement dit correspond à $\theta = 1/2$ et si $\theta = 0$, on retrouve un schéma explicite. Si $\theta = 1$, nous aurons un schéma implicite pur. Les dérivées spatiales s'écrivent (avec un pas constant en x) :

$$\begin{aligned}\left(\frac{\partial u}{\partial x}\right)_j^n &\simeq \frac{U_{j+1}^n - U_{j-1}^n}{2\Delta x} \quad \text{dérivée première centrée en espace} \\ \left(\frac{\partial^2 u}{\partial x^2}\right)_j^n &\simeq \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{\Delta x^2} \quad \text{dérivée seconde centrée en espace}\end{aligned} \tag{11.91}$$

En reportant ces expressions, nous pouvons écrire ce schéma sous la forme générale

$$\alpha_j^+ U_{j+1}^{n+1} + \alpha_j^c U_j^{n+1} + \alpha_j^- U_{j-1}^{n+1} = a_j^+ U_{j+1}^n + a_j^c U_j^n + a_j^- U_{j-1}^n \tag{11.92}$$

avec $\alpha_j^+ = \theta(C/2 - D)$, $\alpha_j^c = 1 + 2\theta D$, $\alpha_j^- = -\theta(D + C/2)$, $a_j^+ = (1-\theta)(D - C/2)$, $a_j^c = 1-2(1-\theta)D$ et $a_j^- = (1-\theta)(D+C/2)$. Ceci conduit à la résolution d'un système d'équations linéaires qui est tridiagonal, soit :

$$\begin{bmatrix} CL & & & & & & \\ \alpha_2^- & \alpha_2^c & \alpha_2^+ & & & & \\ & \ddots & & & & & \\ & \alpha_j^- & \alpha_j^c & \alpha_j^+ & & & \\ & & & \ddots & & & \\ & & & \alpha_{N-1}^- & \alpha_{N-1}^c & \alpha_{N-1}^+ & \\ & & & & & CL & \end{bmatrix} \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_j \\ \vdots \\ U_{N-1} \\ U_N \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_j \\ \vdots \\ d_{N-1} \\ d_N \end{pmatrix} \tag{11.93}$$

En toute rigueur, il faut maintenant étudier la consistance et les conditions de stabilité. La démarche est analogue à celle effectuée dans les paragraphes précédents. Le schéma 11.3 présente la structure de la maille de discrétisation.

$n+1$	$\circ$	$\circ$	$\circ$
$n+1/2$		$+$	
n	$\times$	$\times$	$\times$
	$j-1$	j	$j+1$

figure 11.3 - *Maille de discrétisation du schéma avec $\circ$ point inconnu ; $+$ point d'application de la discrétisation (point fictif), $\times$ point connu*

11.5.3 Stabilité du schéma de Crank-Nicolson

En utilisant la méthode de Von Neumann, il est facile de montrer que le facteur d'amplification g est défini par [2] :

$$g = \frac{a_j^c + (a_j^+ + a_j^-)\cos\phi + i(a_j^+ - a_j^-)\sin\phi}{a_j^c + (a_j^+ + a_j^-)\cos\phi + i(a_j^+ - a_j^-)\sin\phi} \tag{11.94}$$

2. Du fait de la nouvelle définition du paramètre θ dans l'équation (11.87), le paramètre θ de l'équation (11.44) est momentanément remplacé par ϕ

La condition $g^2 < 1$ s'écrit :

$$\begin{aligned}(a_j^+ + a_j^c + a_j^-)^2 - 4z4a_j^+a_j^- + a_j^c(a_j^+ + a_j^-) + 16z^2a_j^+a_j^- < (a_j^+ + a_j^c + a_j^-)^2 \\ - 4z4a_j^+a_j^- + a_j^c(a_j^+ + a_j^-) + 16z^2a_j^+a_j^-\end{aligned} \tag{11.95}$$

En utilisant le fait que, dans le cas de l'équation d'advection diffusion sans terme source, les coefficients vérifient les relations $a_j^+ + a_j^c + a_j^- = a_j^+ + a_j^c + a_j^- = 1$, le module g s'écrit :

$$g = \frac{1-(a_j^+ + a_j^-)(1-\cos\phi) + \mathrm{i}(a_j^+ - a_j^-)\sin\phi}{1-(a_j^+ + a_j^-)(1-\cos\phi) + \mathrm{i}(a_j^+ - a_j^-)\sin\phi} \tag{11.96}$$

Dans ces conditions, en effectuant la même transformation que celle effectuée dans l'étude des schémas explicites, $z = \frac{1-\cos\phi}{2}$, les conditions de stabilité s'écrivent :

$$\begin{aligned}(a_j^+ - a_j^-)^2 - (a_j^+ - a_j^-)^2 - (a_j^+ + a_j^-) + (a_j^+ + a_j^-) \geqslant 0 \\ (a_j^+ + a_j^-)^2 - (a_j^+ + a_j^-)^2 - (a_j^+ + a_j^-) + (a_j^+ + a_j^-) \geqslant 0\end{aligned} \tag{11.97}$$

Ces formules sont tout à fait générales et peuvent être appliquées à tout schéma implicite ou explicite faisant intervenir deux niveaux de temps. Dans le cas du schéma de Crank-Nicolson pondéré, nous avons les deux conditions :

$$(2\theta - 1)\mathrm{C}^2 + 2\mathrm{D} \geqslant 0 \qquad 2\mathrm{D}(1 - 2\mathrm{D}(1-2\theta)) \geqslant 0 \tag{11.98}$$

qui sont toujours vérifiées dès lors que $\theta \geqslant 1/2$. Par conséquent, le schéma de Crank-Nicolson qui correspond à $\theta \geqslant 1/2$ est stable sans condition.

11.5.4 Erreur de troncature du schéma de Crank-Nicolson

Nous allons déterminer l'erreur de troncature du schéma de Crank-Nicolson en effectuant les calculs dans la cas général. Considérons l'équation générale :

$$\alpha_j^+ \mathrm{U}_{j+1}^{n+1} + \alpha_j^c \mathrm{U}_j^{n+1} + \alpha_j^- \mathrm{U}_{j-1}^{n+1} - a_j^+ \mathrm{U}_{j+1}^n - a_j^c \mathrm{U}_j^n - a_j^- \mathrm{U}_{j-1}^n = 0 \tag{11.99}$$

et remplaçons les valeurs discrètes par les valeurs continues puis effectuons le développement en séries de Taylor d'une fonction de deux variables au point $(n+1/2, j)$:

$$\begin{aligned}u_{j+e'}^{n+1/2+e/2} &= u^* + \frac{e\Delta t}{2}u_t^* + e'\Delta x u_x^* \\ &+ \frac{1}{2}\left[\left(\frac{e\Delta t}{2}\right)^2 u_{tt}^* + e\Delta t e'\Delta x u_{tx}^* + (e'\Delta x)^2 u_{xx}^*\right] \\ &+ \frac{1}{3!}\left[\left(\frac{e\Delta t}{2}\right)^3 u_{ttt}^* + 3\left(\frac{e\Delta t}{2}\right)^2 e'\Delta x u_{ttx}^* + 3\frac{e\Delta t}{2}(e'\Delta x)^2 u_{txx}^* + (e'\Delta x)^3 u_{xxx}^*\right] \\ &+ \frac{1}{4!}\left[\left(\frac{e\Delta t}{2}\right)^4 u_{tttt}^* + 4\left(\frac{e\Delta t}{2}\right)^3 e'\Delta x u_{tttx}^*\right. \\ &\qquad \left. + 6\left(\frac{e\Delta t}{2}\right)^2 (e'\Delta x)^2 u_{ttxx}^* 4\left(\frac{e\Delta t}{2}\right)^2 (e'\Delta x)^2 u_{txxx}^* + (e'\Delta x)^4 u_{xxx}^*\right]\end{aligned} \tag{11.100}$$

Les six points du maillage auront les valeurs de (e, e') données dans le tableau 11.1, ce

	$j-1$	j	$j+1$
$n+1$	$(e=1,e'=-1)$	$(e=1,e'=0)$	$(e=1,e'=1)$
n	$(e=-1,e'=-1)$	$(e=-1,e'=0)$	$(e=-1,e'=1)$

tableau 11.1 - *Valeurs aux points du maillage*

qui donne, par exemple pour le point u^n_{j-1}, $e=-1$ et $e'=-1$:

$$\begin{aligned}
u^n_{j-1} &= u^* - \frac{\Delta t}{2}u^*_t - \Delta x u^*_x \\
&+ \frac{1}{2}\left[\left(\frac{\Delta t}{2}\right)^2 u^*_{tt} + \Delta t\Delta x u^*_{tx} + \Delta x^2 u^*_{xx}\right] \\
&+ \frac{1}{3!}\left[\left(-\frac{\Delta t}{2}\right)^3 u^*_{ttt} - 3\left(\frac{\Delta t}{2}\right)^2 \Delta x u^*_{ttx} - 3\frac{\Delta t}{2}\Delta x^2 u^*_{txx} - \Delta x^3 u^*_{xxx}\right] \\
&+ \frac{1}{4!}\left[\left(\frac{\Delta t}{2}\right)^4 u^*_{tttt} + 4\left(\frac{\Delta t}{2}\right)^3 \Delta x u^*_{ttx} + 6\left(\frac{\Delta t}{2}\right)^2 \Delta x^2 u^*_{ttxx}\right. \\
&\qquad \left. + 2\Delta t\Delta x^3 u^*_{txxx} + \Delta x^4 u^*_{xxxx}\right]
\end{aligned} \tag{11.101}$$

et des formes analogues pour les autres points. La forme générale de cette équation est alors :

$$\begin{aligned}
&u^*[\alpha^+_j + a^c_j + a^-_j - (a^+_j + a^c_j + a^-_j)] \\
&+ \frac{\Delta t}{2}u^*_t[a^+_j + a^c_j + a^-_j + a^+_j + a^c_j + a^-_j] + \Delta x u^*_t[a^+_j - a^-_j - a^+_j + a^-_j] \\
&+ \frac{\Delta t^2}{8}u^*_{tt}[a^+_j + a^c_j + a^-_j - (a^+_j + a^c_j + a^-_j)] + \frac{\Delta x}{\Delta t^2}u^*_{tx}[a^+_j - a^-_j + a^+_j - a^-_j] \\
&+ \frac{\Delta x^2}{2}u^*_{xx}[a^+_j + a^-_j - a^+_j + a^-_j] + \frac{\Delta t^3}{48}u^*_{ttt}[a^+_j + a^c_j + a^-_j + a^+_j + a^c_j + a^-_j] \\
&+ \Delta x\frac{\Delta t^2}{8}u^*_{ttx}[a^+_j - a^-_j - (a^+_j - a^-_j)] + \Delta t\frac{\Delta x^2}{4}u^*_{txx}[a^+_j + a^-_j + a^+_j + a^-_j] \\
&+ \frac{\Delta x^3}{6}u^*_{xxx}[a^+_j - a^-_j - (a^+_j + a^-_j)] + \frac{\Delta t^4}{384}u^*_{tttt}[a^+_j + a^c_j + a^-_j - (a^+_j + a^c_j + a^-_j)] \\
&+ \Delta x\frac{\Delta t^3}{48}u^*_{tttx}[a^+_j - a^-_j + a^+_j - a^-_j] + \Delta x^2\frac{\Delta t^2}{16}u^*_{ttxx}[a^+_j - a^-_j - a^+_j - a^-_j] \\
&+ \Delta t\frac{\Delta x^3}{48}u^*_{txxx}[a^+_j - a^-_j + a^+_j - a^-_j] + \frac{\Delta x^4}{24}u^*_{xxxx}[a^+_j + a^-_j - (a^+_j + a^-_j)]
\end{aligned} \tag{11.102}$$

Dans le cas du schéma de Crank-Nicolson, les coefficients du schéma sont :

$$\begin{cases} \alpha^+_j = \mathrm{C}/4 - \mathrm{D}/2 \\ \alpha^c_j = 1 + \mathrm{D} \\ \alpha^-_j = -\mathrm{C}/4 - \mathrm{D}/2 \end{cases} \quad \text{et} \quad \begin{cases} a^+_j = \mathrm{D}/2 - \mathrm{C}/4 \\ a^c_j = 1 - \mathrm{D} \\ a^-_j = \mathrm{D}/2 + \mathrm{C}/4 \end{cases} \tag{11.103}$$

et dans ces conditions, nous avons :

$$\begin{cases} \alpha^+_j + \alpha^c_j + \alpha^-_j - (a^+_j + a^c_j + a^-_j) = 0 \\ \alpha^+_j + \alpha^c_j + \alpha^-_j + (a^+_j + a^c_j + a^-_j) = 2 \\ \alpha^+_j + \alpha^-_j = -\mathrm{D} \end{cases} \quad \text{et} \quad \begin{cases} \alpha^+_j - \alpha^-_j = \mathrm{C}/2 \\ a^+_j + a^-_j = \mathrm{D} \\ a^+_j - a^-_j = -\mathrm{C}/2 \end{cases} \tag{11.104}$$

et le schéma de discrétisation devient :

$$\begin{aligned}&\Delta t\, u_t^* + \mathrm{C}\Delta x\, u_x^* + -\mathrm{D}\Delta x^2 u_{xx}^* + \frac{\Delta t^3}{24} u_{ttt}^* + \frac{\Delta x^3}{6} u_{xxx}^* - \mathrm{D}\Delta t^2 \frac{\Delta x^2}{8} u_{ttxx}^* \\ &- \mathrm{D}\frac{\Delta x^4}{12} u_{xxxx}^* + \mathcal{O}(\Delta t^4) + \mathcal{O}(\Delta x^4) = 0\end{aligned} \tag{11.105}$$

En factorisant Δt, on retrouve l'équation initiale $u_t^* + \mathrm{V}u_x^* - a\,u_{xx}^*$ et des termes d'erreur qui sont en $\mathcal{O}(\Delta t^2)$ et $\mathcal{O}(\Delta x^2)$. Le schéma de Crank-Nicolson est donc un schéma consistant et d'ordre 2 à la fois en temps et en espace.

11.5.5 Schéma implicite pur

Considérons l'équation générale d'un schéma implicite :

$$\alpha_j^+ \mathrm{U}_{j+1}^{n+1} + \alpha_j^c \mathrm{U}_j^{n+1} + \alpha_j^- \mathrm{U}_{j-1}^{n+1} - a_j^+ \mathrm{U}_{j+1}^n - a_j^c \mathrm{U}_j^n - a_j^- \mathrm{U}_{j-1}^n = 0 \tag{11.106}$$

dans laquelle $\theta = 1$; nous obtenons alors le schéma implicite pur qui relient trois valeurs à l'étape inconnue $(n+1)$ à une seule valeur à l'étape connue (n) ; il est facile de vérifier que $a_j^+ = a_j^- = 0$ et que $a_j^c = 1$. Nous obtenons alors l'équation du schéma implicite pur :

$$\alpha_j^+ \mathrm{U}_{j+1}^{n+1} + \alpha_j^c \mathrm{U}_j^{n+1} + \alpha_j^- \mathrm{U}_{j-1}^{n+1} = \mathrm{U}_j^n \tag{11.107}$$

avec $\alpha_j^+ = \mathrm{C}/2 - \mathrm{D}$, $\alpha_j^c = 1 + 2\mathrm{D}$ et $\alpha_j^- = -\mathrm{D} + \mathrm{C}/2$. En utilisant le résultat précédent ; on trouve que les conditions de stabilité sont :

$$\mathrm{C}^2 + 2\mathrm{D} \geqslant 0; \qquad 2\mathrm{D}(1 + 2\mathrm{D}) \geqslant 0 \tag{11.108}$$

conditions qui sont toujours vérifiées pour toutes les valeurs de C et de D.

En conclusion, le schéma implicite pur ou « fully implicit » est stable inconditionnellement ; on peut vérifier cependant que l'erreur de troncature est en $\mathcal{O}(\Delta t)$, donc du premier ordre en temps. Au niveau des calculs, il est clair que l'avantage est faible car l'économie par rapport au schéma de Crank-Nicolson ne réside que dans le calcul du second membre du système tridiagonal, la précision temporelle étant moindre. Une erreur courante est de dire que ce schéma est plus stable que le schéma de Crank-Nicolson : en fait, comme on l'a vu dans les résultats précédents, la stabilité est binaire, soit un schéma est stable, soit il ne l'est pas.

11.5.6 Schéma explicite ADE

Il est possible d'étudier la stabilité du schéma ADE à partir des conditions générales trouvées pour les schémas implicites. Le facteur d'amplification g est défini par :

$$g = \frac{a_j^c + (a_j^+ + a_j^-)\cos\phi + \mathrm{i}(a_j^+ - a_j^-)\sin\phi}{\alpha_j^c + (\alpha_j^+ + \alpha_j^-)\cos\phi + \mathrm{i}(\alpha_j^+ - \alpha_j^-)\sin\phi} \tag{11.109}$$

et on doit vérifier la condition (11.110) pour $0 \leqslant z \leqslant 1$:

$$\begin{aligned}(a_j^+ + a_j^c + a_j^-)^2 - 4z4a_j^+ a_j^- + a_j^c(a_j^+ + a_j^-) + 16z^2 a_j^+ a_j^- \\ < (\alpha_j^+ + \alpha_j^c + \alpha_j^-)^2 - 4z4\alpha_j^+ \alpha_j^- + \alpha_j^c(\alpha_j^+ + \alpha_j^-) + 16z^2 \alpha_j^+ \alpha_j^-\end{aligned} \tag{11.110}$$

Dans le cas de la solution P, nous avons $\alpha_j^+ = 0$, $\alpha_j^c = 1+D$, $\alpha_j^- = -(D+C/2)$, $a_j^+ = D-C/2$, $a_j^c = 1-D$ et $a_j^- = 0$. On vérifie que $(a_j^+ + a_j^c + a_j^-)^2 = (\alpha_j^+ + \alpha_j^c + \alpha_j^-)^2$ et l'inégalité se simplifie. $a_j^c a_j^+ > \alpha_j^c \alpha_j^-$, soit $D-C/2-D^2+DC/2 > -D-C/2-D^2-DC/2$, ce qui conduit à la condition :

$$D(2+C) > 0 \tag{11.111}$$

Pour la solution Q, nous obtenons $\alpha_j^+ = C/2-D$, $\alpha_j^c = 1+D$, $\alpha_j^- = 0$, $a_j^+ = 0$, $a_j^c = 1-D$, $a_j^- = D+C/2$ et la condition de stabilité s'écrit alors $a_j^c a_j^- > \alpha_j^c \alpha_j^+$, soit $D+C/2-D^2-DC/2 > C/2-D+DC/2-D^2$, ce qui conduit à la condition :

$$D(2-C) > 0 \tag{11.112}$$

Comme le nombre de diffusion est positif, les deux conditions se traduisent par :

$$|C| = \frac{|V|\Delta t}{\Delta x} < 2 \tag{11.113}$$

Si le pas d'espace Δx est fixé, le pas de temps Δt devra vérifier la condition :

$$\Delta t < \frac{2\Delta x}{|V|} \tag{11.114}$$

On voit que dans le cas d'un problème diffusif pur, la méthode ADE est inconditionnellement stable. Par contre, dans le cas où il y a de la convection, la limitation du pas de temps est deux fois plus grande que dans le cas d'un schéma explicite FTCS. Par contre, la précision temporelle est en $\mathcal{O}(\Delta t^2)$.

11.6 Schémas à deux pas de temps

11.6.1 Problème

On considère un schéma de discrétisation faisant intervenir des points aux temps $t = t^n + \Delta t$, $t = t^n$ et $t = t^n - \Delta t$ et faisant intervenir les valeurs de u aux positions x, x_{j-1}, x_j et x_{j+1}. Le fait de faire intervenir trois niveaux de temps ($n+1$, n et $n-1$) devrait permettre d'écrire au point (n, j) un schéma explicite précis au second ordre en temps.

11.6.2 Schéma de Richardson

Ce schéma est obtenu en écrivant l'opérateur différentiel en un point correspondant au temps $t = t^n + \Delta t$ et à l'abscisse $x = x_j$. La dérivée temporelle $\frac{\partial u}{\partial t}$ s'écrit :

$$\left(\frac{\partial u}{\partial t}\right)_j^n \simeq \frac{U_j^{n+1} - U_j^{n-1}}{2\Delta x} \quad \text{dérivée centrée en temps} \tag{11.115}$$

et les dérivées spatiales sont évaluées au temps n, soit (avec un pas constant Δx en x) :

$$\left(\frac{\partial u}{\partial x}\right)_j^n \simeq \frac{U_{j+1}^n - U_{j-1}^n}{2\Delta x} \quad \text{dérivée première centrée en espace} \tag{11.116}$$

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_j^n \simeq \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{\Delta x^2} \quad \text{dérivée seconde centrée en espace} \tag{11.117}$$

En reportant ces expressions, nous pouvons écrire le schéma sous la forme suivante

$$\frac{U_j^{n+1}-U_j^{n-1}}{2\Delta t}=\frac{-VU_{j+1}^n-U_{j-1}^n}{2\Delta x}+\frac{aU_{j+1}^n-2U_j^n+U_{j-1}^n}{\Delta x^2} \tag{11.118}$$

ou sous forme générale :

$$U_j^{n+1}=U_j^{n-1}+a_j^+U_{j+1}^n+a_j^cU_j^n+a_j^-U_{j-1}^n \tag{11.119}$$

avec $a_j^+=2D-C$, $a_j^c=-4D$ et $a_j^-=2D+C$ en introduisant le nombre de Courant et le nombre de diffusion $C=\frac{V\Delta t}{\Delta x}$ et $D=\frac{a\Delta t}{\Delta x^2}$. Le schéma 11.4 présente la structure de la maille de discrétisation. On peut montrer facilement que ce schéma est du second ordre en espace et en temps.

$n+1$		$\circ$	
n	$\times$	$+$	$\times$
$n-1$		$\times$	
	$j-1$	j	$j+1$

figure 11.4 - *Maille de discrétisation du schéma avec* $\circ$ *point inconnu ;* $+$ *point d'application de la discrétisation (point fictif),* $\times$ *point connu*

11.6.3 Stabilité du schéma de Richardson

En utilisant la méthode de Von Neumann ; il est facile de monter que l'erreur se met sous la forme :

$$E^{n+1}=E^n(a_j^+e^{i\theta}+a_j^c+a_j^-e^{-i\theta})+E^{n-1} \tag{11.120}$$

Dans ces conditions, si on appelle g le facteur d'amplification de l'erreur entre deux pas de temps consécutifs, nous aurons $E^n=gE^{n-1}$ et $E^{n+1}=gE^n=g^2E^{n-1}$. En remplaçant dans l'expression ci-dessus, il vient :

$$g^2-g(a_j^+e^{i\theta}+a_j^c+a_j^-e^{-i\theta})-1=0 \tag{11.121}$$

et il faut déterminer les conditions pour que g, racine de cette équation ; soit de module inférieur ou égal à 1. Pour résoudre ce problème on utilise le théorème de Miller (1971).

Une autre possibilité est de travailler sur la matrice d'amplification **G** définie par :

$$\mathbf{G}=\begin{bmatrix} a_j^+e^{i\theta}+a_j^c+a_j^-e^{-i\theta} & 1 \\ 1 & 0 \end{bmatrix} \tag{11.122}$$

et la condition est que le rayon spectral de cette matrice soit inférieur à 1.

Matrice d'amplification

Si on écrit la matrice d'amplification **G** sous la forme :

$$\mathbf{G}=\begin{bmatrix} r & 1 \\ 1 & 0 \end{bmatrix} \tag{11.123}$$

le polynôme caractéristique est $P(\lambda) = \lambda^2 - \lambda r - 1$. Le discriminant est $\Delta = r^2 + 4$ et les deux racines sont $\lambda = \frac{-r \pm \Delta^{1/2}}{2}$. Il faut donc étudier le module de chacune de ces racines. Pour le schéma de Richardson, $r = 4D(\cos\theta - 1) - 4iC\sin\theta$; si on ne considère que le cas diffusif pur, $r = 4D(\cos\theta - 1)$ et $D = 1 + 4D^2(\cos\theta - 1)^2$ étant positif, r étant négatif, la racine de plus grand module est $\lambda = -r - \Delta^{1/2} = 4D(1 - \cos\theta) + (1 + 4D^2(\cos\theta - 1)^2)^{1/2}$ qui est de module supérieur ou égal à 1 ; donc le schéma de Richardson est *inconditionnellement instable*.

Coefficient d'amplification : théorème de Miller

Soit $f(\lambda)$ un polynôme à coefficients complexes de degré n donné sous la forme :

$$f(\lambda) = \sum_{i=0}^{n} a_i \lambda^i \tag{11.124}$$

si on définit le polynôme f' par :

$$f'(\lambda) = \sum_{i=0}^{n} a_i^* \lambda^{n-i} \tag{11.125}$$

où a_i^* est le conjugué de a_i, le polynôme $f_1(l)$ défini par :

$$f_1(\lambda) = \frac{1}{\lambda}(f'(0)f(\lambda) - f(0)f'(\lambda)) \tag{11.126}$$

est de degré inférieur à n. Les racines du polynôme $f(?)$ sont de module inférieur ou égal à 1 si et seulement si :

(i) si $|f'(0)| > |f(0)|$ les racines de $f_1(\lambda) = 0$ sont de module inférieur ou égal à 1 ;

(ii) si f_1 est nul les racines de $df/d\lambda = 0$ sont de module inférieur ou égal à 1.

Application au schéma de Richardson

En remplaçant les coefficients $a^{+,c,-}$ par leurs valeurs, il vient :

$$\begin{aligned} &g^2 + g[4D(\cos\theta - 1) - 4iC\sin\theta] - 1 = 0 \\ &f(g) = g^2 + g[4D(\cos\theta - 1) - 4iC\sin\theta] - 1 \\ &f'(g) = 1 + g[4D(\cos\theta - 1) + 4iC\sin\theta] - g^2 \\ &f(0) = -1 \\ &f'(0) = 1 \end{aligned} \tag{11.127}$$

et :

$$f_1(g) = [f(g) + f'(g)]/g = 8D(\cos\theta - 1) \tag{11.128}$$

et ce polynôme n'a pas de racine de module inférieur à 1, donc le schéma est inconditionnellement instable ; en clair, ce schéma ne marche jamais.

11.6.4 Schéma de Dufort et Frankel

Ce schéma est obtenu comme dans le schéma précédent en écrivant l'opérateur différentiel en un point correspondant au temps $t = t^n + \Delta t$ et à l'abscisse $x = x_j$ mais en modifiant l'écriture de la dérivée seconde afin de modifier la matrice d'amplification G

et par suite ses valeurs propres ; le but étant de trouver un schéma stable. La dérivée temporelle $\frac{\partial u}{\partial t}$ s'écrit :

$$\left(\frac{\partial u}{\partial t}\right)_j^n \simeq \frac{U_j^{n+1} - U_j^{n-1}}{2\Delta t} \quad \text{dérivée centrée en temps} \tag{11.129}$$

et les dérivées spatiales sont évaluées au temps n, soit (avec un pas constant Δx en x) :

$$\begin{aligned} \left(\frac{\partial u}{\partial x}\right)_j^n &\simeq \frac{U_{j+1}^n - U_{j-1}^n}{2\Delta x} \quad \text{dérivée première centrée en espace} \\ \left(\frac{\partial^2 u}{\partial x^2}\right)_j^n &\simeq \frac{U_{j+1}^n - 2\left(\theta U_j^{n+1} + (1-\theta)U_j^{n-1}\right) + U_{j-1}^n}{2\Delta x^2} \quad \begin{array}{l}\text{dérivée seconde} \\ \text{en espace}\end{array} \end{aligned} \tag{11.130}$$

Le schéma de Dufort et Frankel s'écrit alors sous la forme :

$$\frac{U_j^{n+1} - U_j^{n-1}}{2\Delta t} = -\frac{VU_{j+1}^n + U_{j-1}^n}{2\Delta x} + \frac{aU_{j+1}^n - 2\left(\theta U_j^{n+1} + (1-\theta)U_j^{n-1}\right) + U_{j-1}^n}{\Delta x^2} \tag{11.131}$$

et on peut remarquer que, contrairement au schéma de Richardson, le schéma de Dufort et Frankel ne fait pas intervenir la valeur « centrale » U_j^n mais une pondération entre les deux valeurs aux temps $(n-1)\Delta t$ et $(n+1)\Delta t$.

Le schéma 11.5 présente la structure de la maille. On peut monter que si ce schéma est stable par contre, il n'est pas forcément consistant.

$n+1$		$\circ$	
n	$\times$	$+$	$\times$
$n-1$		$\times$	
	$j-1$	j	$j+1$

figure 11.5 - *Maille de discrétisation du schéma avec* $\circ$ *point inconnu ;* $+$ *point d'application de la discrétisation (point fictif),* $\times$ *point connu*

11.7 Problèmes à deux dimensions d'espace

11.7.1 Formulation

On considère une équation aux dérivées partielles du second ordre de type parabolique à trois variables, t, x et y. Soit $u(x,y,t)$ cette fonction et l'opérateur différentiel $\mathscr{L}$ est défini par :

$$\mathscr{L}(u) = \frac{\partial u}{\partial t} + V_x\frac{\partial u}{\partial x} + V_y\frac{\partial u}{\partial y} - a_x\frac{\partial^2 u}{\partial x^2} - a_y\frac{\partial^2 u}{\partial^2 y} \tag{11.132}$$

que l'on peut mettre sous la forme :

$$\mathscr{L}(u) = \frac{\partial u}{\partial t} + \mathscr{L}_x(u) + \mathscr{L}_y(u) \tag{11.133}$$

où $\mathscr{L}_x(u)$ et $\mathscr{L}_y(u)$ sont deux opérateurs différentiels unidirectionnels dans les deux directions x et y. L'opérateur différentiel précédent correspond à une équation d'advection diffusion dans laquelle les coefficients de diffusion sont constants :

$$\mathscr{L}_x = V_x\frac{\partial}{\partial x} - a_x\frac{\partial^2}{\partial x^2} \quad \text{et} \quad \mathscr{L}_y = V_y\frac{\partial}{\partial y} - a_y\frac{\partial^2}{\partial y^2} \tag{11.134}$$

a_x et a_y sont des coefficients de diffusion constants dans chacune des deux directions x et y respectivement ; si la diffusion est isotrope, $a_x = a_y = a$. Lorsque les coefficients de diffusion sont variables, l'opérateur différentiel est :

$$\mathscr{L}(u) = \frac{\partial u}{\partial t} + \mathrm{V}_x \frac{\partial u}{\partial x} + \mathrm{V}_y \frac{\partial u}{\partial y} - \frac{\partial}{\partial x}\left(a_x \frac{\partial u}{\partial x}\right) - \frac{\partial}{\partial y}\left(a_y \frac{\partial u}{\partial y}\right) \tag{11.135}$$

et les deux opérateurs spatiaux sont :

$$\mathscr{L}_x = \mathrm{V}_x \frac{\partial}{\partial x} - \frac{\partial}{\partial x}\left(a_x \frac{\partial}{\partial x}\right) \quad \text{et} \quad \mathscr{L}_y = \mathrm{V}_y \frac{\partial}{\partial y} - \frac{\partial}{\partial y}\left(a_y \frac{\partial}{\partial y}\right) \tag{11.136}$$

Soit à résoudre le problème suivant :

$$\begin{aligned} &\mathscr{L}(u) = 0 \qquad \text{pour} \quad t > 0 \quad \text{et} \quad (x,y) \in]x_a \;\; x_b[\times]y_a \;\; y_b[\\ &u(0,x,y) \text{ donné (condition initiale)} \\ &u(t,x_a,y), u(t,x_b,y), u(t,x,y_a) \text{ et } u(t,x,y_b) \text{ donnés (conditions limites)} \end{aligned} \tag{11.137}$$

Le domaine de calcul $t > 0, x_a \leqslant x \leqslant x_b, y_a \leqslant y \leqslant y_b$ est discrétisé comme précédemment et on notera $\mathrm{U}_{i,j}^n \simeq u(t^n, x_i, y_j)$ la valeur discrète approchée de la fonction u au temps t^n et au point d'abscisse x_i et d'ordonnée y_j.

11.7.2 Schéma explicite FTCS

Ce schéma (Forward Time Centered Space) est obtenu en écrivant l'opérateur différentiel au point correspondant au temps $t = t^n$ et au point d'abscisse $x = x_i$ et d'ordonnée $y = y_j$. La dérivée temporelle s'écrit :

$$\left(\frac{\partial u}{\partial t}\right)_j^n \simeq \frac{\mathrm{U}_{i,j}^{n+1} - \mathrm{U}_{i,j}^{n-1}}{\Delta t} \quad \text{dérivée avant en temps} \tag{11.138}$$

alors que les dérivées spatiales s'écrivent (avec un pas constant en x et y) :

$$\left(\frac{\partial u}{\partial x}\right)_j^n \simeq \frac{\mathrm{U}_{i+1,j}^n - \mathrm{U}_{i-1,j}^n}{2\Delta x} \quad \begin{array}{l}\text{dérivée première par rapport à } x \\ \text{centrée en espace}\end{array} \tag{11.139}$$

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_j^n \simeq \frac{\mathrm{U}_{i+1,j}^n - 2\mathrm{U}_{i,j}^n + \mathrm{U}_{i-1,j}^n}{\Delta x^2} \quad \begin{array}{l}\text{dérivée seconde par rapport à } x \\ \text{centrée en espace}\end{array} \tag{11.140}$$

$$\left(\frac{\partial u}{\partial y}\right)_j^n \simeq \frac{\mathrm{U}_{i,j+1}^n - \mathrm{U}_{i,j-1}^n}{2\Delta y} \quad \begin{array}{l}\text{dérivée première par rapport à } y \\ \text{centrée en espace}\end{array} \tag{11.141}$$

$$\left(\frac{\partial^2 u}{\partial y^2}\right)_j^n \simeq \frac{\mathrm{U}_{i,j+1}^n - 2\mathrm{U}_{i,j}^n + \mathrm{U}_{i,j-1}^n}{\Delta y^2} \quad \begin{array}{l}\text{dérivée seconde par rapport à } y \\ \text{centrée en espace}\end{array} \tag{11.142}$$

En remplaçant ces expressions dans l'équation aux dérivées partielles à résoudre, il vient :

$$\begin{aligned} \frac{1}{\Delta t}\left(\mathrm{U}_{i,j}^{n+1} - \mathrm{U}_{i,j}^n\right) = &-\frac{\mathrm{V}_x}{2\Delta x}\left(\mathrm{U}_{i+1,j}^n - \mathrm{U}_{i-1,j}^n\right) + \frac{a}{\Delta x^2}\left(\mathrm{U}_{i+1,j}^n - 2\mathrm{U}_{i,j}^n + \mathrm{U}_{i-1,j}^n\right) \\ &-\frac{\mathrm{V}_y}{2\Delta y}\left(\mathrm{U}_{i,j+1}^n - \mathrm{U}_{i,j-1}^n\right) + \frac{a}{\Delta y^2}\left(\mathrm{U}_{i,j+1}^n - 2\mathrm{U}_{i,j}^n + \mathrm{U}_{i,j-1}^n\right) \end{aligned} \tag{11.143}$$

soit, en explicitant :

$$\begin{aligned} \mathrm{U}_{i,j}^{n+1} = & \mathrm{U}_{i+1,j}^{n}\left(\frac{a\Delta t}{\Delta x^2}-\frac{\mathrm{V}_x\Delta t}{2\Delta x}\right)+\mathrm{U}_{i,j}^{n}\left(1-2\frac{a\Delta t}{\Delta x^2}-2\frac{a\Delta t}{\Delta y^2}\right) \\ & +\mathrm{U}_{i-1,j}^{n}\left(\frac{a\Delta t}{\Delta x^2}+\frac{\mathrm{V}_x\Delta t}{2\Delta x}\right)+\mathrm{U}_{i,j+1}^{n}\left(\frac{a\Delta t}{\Delta y^2}-\frac{\mathrm{V}_y\Delta t}{2\Delta y}\right) \\ & +\mathrm{U}_{i,j-1}^{n}\left(\frac{a\Delta t}{\Delta y^2}+\frac{\mathrm{V}_y\Delta t}{2\Delta y}\right) \end{aligned} \tag{11.144}$$

Si les valeurs des deux pas de discrétisation en espace ne sont pas identiques, nous aurons deux nombres de Courant et deux nombres de diffusion définis respectivement par $\mathrm{C}_x=\frac{\Delta t\mathrm{V}_x}{\Delta x}$, $\mathrm{C}_y=\frac{\Delta t\mathrm{V}_y}{\Delta y}$, $\mathrm{D}_x=\frac{a\Delta t}{\Delta x^2}$ et $\mathrm{D}_y=\frac{a\Delta t}{\Delta y^2}$. Avec ces notations, si on définit les points voisins du point (i,j) par un repérage géographique (nord, sud, est, ouest) :

$$\mathrm{U}_{i,j}^{n+1}=a_{\mathrm{N}}^{n}\mathrm{U}_{i,j+1}^{n}+a_{\mathrm{S}}^{n}\mathrm{U}_{i,j-1}^{n}+a_{\mathrm{W}}^{n}\mathrm{U}_{i-1,j}^{n}+a_{\mathrm{E}}^{n}\mathrm{U}_{i+1,j}^{n}+a_{\mathrm{C}}^{n}\mathrm{U}_{i,j}^{n} \tag{11.145}$$

avec $a_{\mathrm{E}}^{n}=(\frac{a\Delta t}{\Delta x^2}-\frac{\Delta t\mathrm{V}_x}{2\Delta x})$, $a_{\mathrm{C}}^{n}=(1-2\frac{a\Delta t}{\Delta x^2}-2\frac{a\Delta t}{\Delta y^2})$, $a_{\mathrm{W}}^{n}=(\frac{a\Delta t}{\Delta x^2}+\frac{\Delta t\mathrm{V}_x}{2\Delta x})$, $a_{\mathrm{N}}^{n}=(\frac{a\Delta t}{\Delta y^2}-\frac{\Delta t\mathrm{V}_y}{2\Delta y})$ et $a_{\mathrm{S}}^{n}=(\frac{a\Delta t}{\Delta y^2}+\frac{\Delta t\mathrm{V}_y}{2\Delta y})$ ou $a_{\mathrm{E}}^{n}=\mathrm{D}_x-\mathrm{C}_x/2$, $a_{\mathrm{C}}^{n}=1-2\mathrm{D}_x-2\mathrm{D}_y$, $a_{\mathrm{W}}^{n}=\mathrm{D}_x+\mathrm{C}_x/2$, $a_{\mathrm{N}}^{n}=\mathrm{D}_y-\mathrm{C}_y/2$ et $a_{\mathrm{S}}^{n}=\mathrm{D}_y+\mathrm{C}_y/2$.

11.7.3 Méthodes à directions alternées ADI

Cette méthode (Alternative Direction Implicit) est basée sur la décomposition de l'opérateur différentiel en deux. Deux demi pas de temps successifs sont considérés et à chacun d'entre eux, on résoudra dans une direction en utilisant un schéma implicite.

$$\mathscr{L}(u)=\frac{\partial u}{\partial t}+\mathscr{L}_x(u)+\mathscr{L}_y(u) \tag{11.146}$$

(i) premier demi pas (au temps $t^{n+\frac{1}{2}}$) :

$$\left(\frac{\partial u}{\partial t}\right)^{n+1/2}=-\mathscr{L}_x^{n+1/2}(u)-\mathscr{L}_y^{n}(u) \tag{11.147}$$

soit :

$$\left(\frac{\partial u}{\partial t}\right)^{n+1/2}+\mathscr{L}_x^{n+1/2}(u)=-\mathscr{L}_y^{n}(u) \tag{11.148}$$

(ii) deuxième demi pas (au temps t^{n+1}) :

$$\left(\frac{\partial u}{\partial t}\right)^{n+1/2}=-\mathscr{L}_x^{n+1/2}(u)-\mathscr{L}_y^{n+1}(u) \tag{11.149}$$

soit :

$$\left(\frac{\partial u}{\partial t}\right)^{n+1/2}+\mathscr{L}_y^{n+1}(u)=-\mathscr{L}_x^{n+1/2}(u) \tag{11.150}$$

au premier demi pas, les dérivées par rapport à y sont évaluées au temps t^n donc à une étape connue ; de même au deuxième demi pas, les dérivées par rapport à x sont évaluées au temps $t^{n+1/2}$ étape qui vient d'être calculée.

Explicitons ces deux étapes :

(i) premier demi pas :

$$\begin{aligned}\frac{\mathrm{U}_{i,j}^{n+1/2}}{\Delta t}+\mathrm{V}_x\frac{\mathrm{U}_{i+1,j}^{n+1/2}-\mathrm{U}_{i-1,j}^{n+1/2}}{2\Delta x}-a\frac{\mathrm{U}_{i+1,j}^{n+1/2}-2\mathrm{U}_{i,j}^{n+1/2}+\mathrm{U}_{i-1,j}^{n+1/2}}{\Delta x^2}\\ =\frac{\mathrm{U}_{i,j}^{n}}{\Delta t}-\mathrm{V}_y\frac{\mathrm{U}_{i,j+1}^{n}-\mathrm{U}_{i,j-1}^{n}}{2\Delta y}+a\frac{\mathrm{U}_{i,j+1}^{n}-2\mathrm{U}_{i,j}^{n}+\mathrm{U}_{i,j-1}^{n}}{\Delta y^2}\end{aligned} \tag{11.151}$$

On obtient donc, à j fixé, un système tridiagonal reliant les inconnues (en i) au temps $t^{n+1/2}$:

$$\begin{aligned}&\mathrm{U}_{i+1,j}^{n+1/2}\left(\frac{\mathrm{V}_x}{2\Delta x}-\frac{a}{\Delta x^2}\right)+\mathrm{U}_{i,j}^{n+1/2}\left(\frac{1}{\Delta t}+\frac{2a}{\Delta x^2}\right)-\mathrm{U}_{i-1,j}^{n+1/2}\left(\frac{\mathrm{V}_x}{2\Delta x}+\frac{a}{\Delta x^2}\right)\\ &\quad=\mathrm{U}_{i,j+1}^{n}\left(-\frac{\mathrm{V}_y}{2\Delta y}+\frac{a}{\Delta y^2}\right)+\mathrm{U}_{i,j}^{n}\left(\frac{1}{\Delta t}+\frac{2a}{\Delta x^2}\right)+\mathrm{U}_{i,j-1}^{n}\left(\frac{\mathrm{V}_y}{2\Delta y}+\frac{a}{\Delta y^2}\right)\end{aligned} \tag{11.152}$$

(ii) deuxième demi pas :

$$\begin{aligned}\frac{\mathrm{U}_{i,j}^{n+1}}{\Delta t}+\mathrm{V}_y\frac{\mathrm{U}_{i,j+1}^{n+1}-\mathrm{U}_{i,j-1}^{n+1}}{2\Delta y}-a\frac{\mathrm{U}_{i,j+1}^{n+1}-2\mathrm{U}_{i,j}^{n+1}+\mathrm{U}_{i,j-1}^{n+1}}{\Delta y^2}\\ =\frac{\mathrm{U}_{i,j}^{n+1/2}}{\Delta t}-\mathrm{V}_x\frac{\mathrm{U}_{i+1,j}^{n+1/2}-\mathrm{U}_{i-1,j}^{n+1/2}}{2\Delta x}+a\frac{\mathrm{U}_{i+1,j}^{n+1/2}-2\mathrm{U}_{i,j}^{n+1/2}+\mathrm{U}_{i-1,j}^{n+1/2}}{\Delta x^2}\end{aligned} \tag{11.153}$$

à i fixé, nous obtenons aussi un système tridiagonal qui relient les inconnues en j au temps t_{n+1}. En faisant le bilan des opérations, si M est le nombre de points dans la direction i et N, le nombre dans la direction j, sont à résoudre au premier demi pas $\mathrm{N}-2$ systèmes tridiagonaux (de M équations à M inconnues) et au deuxième, $\mathrm{M}-2$ systèmes de N équations à N inconnues, soit au total $\mathrm{M}+\mathrm{N}-4$ systèmes.

Cas d'un terme « source »

Dans de nombreux problèmes, il existe un terme complémentaire dans l'équation de transport ; ce terme est supposé, dans un premier temps, indépendant de la variable u. L'équation s'écrit alors :

$$\mathscr{L}(u)=\frac{\partial u}{\partial t}+\mathscr{L}_x(u)+\mathscr{L}_y(u)+\mathrm{S} \tag{11.154}$$

(i) premier demi pas (au temps $t^{n+1/2}$)

$$\left(\frac{\partial u}{\partial t}\right)^{n+1/2}=-\mathscr{L}_x^{n+1/2}(u)-\mathscr{L}_y^{n}(u)-\mathrm{S}^{n+1/2} \tag{11.155}$$

(ii) deuxième demi pas (au temps t^{n+1}) :

$$\left(\frac{\partial u}{\partial t}\right)^{n+1/2}=-\mathscr{L}_x^{n+1/2}(u)-\mathscr{L}_y^{n+1}(u)-\mathrm{S}^{n+1/2} \tag{11.156}$$

Explicitons ces deux étapes :

(i) premier demi pas :

$$\begin{aligned}\frac{\mathrm{U}_{i,j}^{n+1/2}}{\Delta t}+\mathrm{V}_x\frac{\mathrm{U}_{i+1,j}^{n+1/2}-\mathrm{U}_{i-1,j}^{n+1/2}}{2\Delta x}-a\frac{\mathrm{U}_{i+1,j}^{n+1/2}-2\mathrm{U}_{i,j}^{n+1/2}+\mathrm{U}_{i-1,j}^{n+1/2}}{\Delta x^2}\\ =\frac{\mathrm{U}_{i,j}^{n}}{\Delta t}-\mathrm{V}_y\frac{\mathrm{U}_{i,j+1}^{n}-\mathrm{U}_{i,j-1}^{n}}{2\Delta y}+a\frac{\mathrm{U}_{i,j+1}^{n}-2\mathrm{U}_{i,j}^{n}+\mathrm{U}_{i,j-1}^{n}}{\Delta y^2}-\mathrm{S}^{n+1/2}\end{aligned} \tag{11.157}$$

Il s'agit encore à j fixé, d'un système tridiagonal reliant les inconnues (en i) au temps $t^{n+1/2}$.

(ii) deuxième demi pas :

$$\begin{aligned}\frac{\mathrm{U}_{i,j}^{n+1}}{\Delta t}+\mathrm{V}_y\frac{\mathrm{U}_{i,j+1}^{n+1}-\mathrm{U}_{i,j-1}^{n+1}}{2\Delta y}-a\frac{\mathrm{U}_{i,j+1}^{n+1}-2\mathrm{U}_{i,j}^{n+1}+\mathrm{U}_{i,j-1}^{n+1}}{\Delta y^2}=\frac{\mathrm{U}_{i,j}^{n+1/2}}{\Delta t}\\ -\mathrm{V}_x\frac{\mathrm{U}_{i+1,j}^{n+1/2}-\mathrm{U}_{i-1,j}^{n+1/2}}{2\Delta x}+a\frac{\mathrm{U}_{i+1,j}^{n+1/2}-2\mathrm{U}_{i,j}^{n+1/2}+\mathrm{U}_{i-1,j}^{n+1/2}}{\Delta x^2}-\mathrm{S}^{n+1/2}\end{aligned} \tag{11.158}$$

11.8 Dispersion et dissipation numériques

11.8.1 Problème

Dans de nombreux cas, il est utile de caractériser certains effets particuliers du schéma numérique qui est utilisé. Dans les paragraphes précédents, on a montré les phénomènes d'oscillations des solutions et mis en évidence des conditions de stabilité. On considère une équation aux dérivées partielles d'ordre 1 à deux variables (x,t) qui correspond à la convection linéaire d'une solution initiale. Considérons le problème suivant :

$$\begin{aligned}&\frac{\partial u}{\partial t}+\mathrm{V}\frac{\partial u}{\partial x}=0 \qquad \text{pour } t>0 \text{ et } x\in[0\ 1]\\ &u(x,0)=\begin{cases}\sin(a\pi x) & \text{pour } x\in[0\ 1/a]\\ 0 & \text{pour } x\in[1/a\ 1]\end{cases}\\ &u(0,t)=u(1,t)=0\end{aligned} \tag{11.159}$$

La solution analytique du problème pour $t=\frac{a-1}{a\mathrm{V}}$ est :

$$u(x,t)=\begin{cases}0 & \text{si } x\in[0\ \ \mathrm{V}t]\\ \sin(a\pi(x-\mathrm{V}t)) & \text{si } x\in[\mathrm{V}t\ \ \mathrm{V}t+1/a]\\ 0 & \text{si } x[\mathrm{V}t+1/a\ \ 1]\end{cases} \tag{11.160}$$

La solution analytique se propage donc sans que son amplitude soit modifiée. Dans la pratique, de nombreux schémas vont propager la condition initiale mais en diminuant la valeur de l'amplitude et la vitesse de propagation obtenue numériquement sera plus faible que la valeur réelle V ; de plus, le signal obtenu numériquement fait apparaître des longueurs d'onde différentes de la longueur d'onde théorique. On obtient ainsi une solution amortie (dissipation) avec des longueurs d'onde différentes (dispersion). Dans l'exemple choisi, la condition initiale comprenait une seule longueur d'onde, on peut imaginer que si la condition initiale est développée en séries de Fourier, chaque composante sera modifiée comme l'onde étudiée.

On peut étendre ce concept au cas d'une équation aux dérivées partielles d'ordre 2.

11.8.2 Dispersion - dissipation

Examinons deux équations aux dérivées partielles proches de l'équation de convection étudiée au paragraphe précédent :

$$\begin{aligned}\frac{\partial u}{\partial t}+\mathrm{V}\frac{\partial u}{\partial x}-a\frac{\partial^2 u}{\partial x^2}&=0\\ \frac{\partial u}{\partial t}+\mathrm{V}\frac{\partial u}{\partial x}+b\frac{\partial^3 u}{\partial x^3}&=0\end{aligned} \tag{11.161}$$

la première est l'équation d'advection diffusion étudiée dans ce chapitre et la seconde est l'équation linéaire de Korteweg de Vries. L'équation générale de la propagation d'une onde plane amortie est de la forme :

$$u(x,t)=\mathrm{U}e^{-pt}e^{-\mathrm{i}m(x-qt)} \tag{11.162}$$

où p et q sont des fonctions du nombre d'onde m associé à la longueur d'onde $\lambda = 2\pi p/m$ caractérise l'atténuation de l'amplitude U et q est la vitesse de propagation de l'onde.

Pour trouver les valeurs de p et q, il faut satisfaire une des deux équations aux dérivées partielles, ainsi, pour la première, nous avons $p = am^2$ et $q = \mathrm{V}$ alors que pour la seconde, nous avons $p = 0$ et $q = \mathrm{V} - bm^2$. On peut vérifier facilement que pour l'équation de convection pure, nous avons $p = 0$ et $q = \mathrm{V}$.

Ainsi, si on considère l'équation d'advection diffusion ; on constate que l'amplitude des ondes est atténuée par la présence des termes de diffusion ($a\frac{\partial^2 u}{\partial x^2}$) et que p varie quadratiquement avec m, mais que la vitesse de propagation est conservée. On peut remarquer que les longueurs d'onde courtes (nombre d'onde m associé élevé) seront plus amorties que les longueurs d'ondes élevées (nombre d'onde m associé faible). Par contre, avec l'équation de Korteweg de Vries, l'amplitude est conservée, mais la vitesse de propagation de l'onde est modifiée et ce en fonction de sa propre longueur d'onde ; ainsi, au lieu d'avoir une vitesse de propagation constante et égale à V, chaque longueur d'onde sera propagée à sa propre vitesse q. Il y aura donc dispersion.

Lorsque l'on discrétise une équation de convection pure, il faut regarder quels sont les termes négligés et l'erreur de troncature fournit les renseignements sur le comportement du schéma. Ainsi, si les termes prépondérants sont en u'', alors le schéma sera diffusif alors que si les termes sont en u''', alors le schéma sera dispersif.

11.8.3 Dissipation numérique et stabilité de Hirt

Hirt (1968) a proposé une méthode d'analyse de stabilité utilisant l'équation discrétisée en y intégrant les termes de l'erreur de troncature. Si on reprend le schéma FTCS et l'étude de la consistance, nous avons :

$$\frac{\partial u}{\partial t}+\mathrm{V}\frac{\partial u}{\partial x}-a\frac{\partial^2 u}{\partial x^2}=-\frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2}+\mathcal{O}(\Delta t^2,\Delta x^2) \tag{11.163}$$

et cette équation peut être réécrite sous la forme :

$$\frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2}+\frac{\partial u}{\partial t}+\mathrm{V}\frac{\partial u}{\partial x}-a\frac{\partial^2 u}{\partial x^2}=0 \tag{11.164}$$

ou encore :

$$\frac{\Delta t}{2a}\frac{\partial^2 u}{\partial t^2}+\frac{1}{a}\frac{\partial u}{\partial t}+\frac{\mathrm{V}}{a}\frac{\partial u}{\partial x}-\frac{\partial^2 u}{\partial x^2}=0 \tag{11.165}$$

Or du fait de la présence du terme $\frac{\partial^2 u}{\partial t^2}$, la nature de l'équation aux dérivées partielles, qui était parabolique, change et devient hyperbolique. Les conditions de stabilité (Courant Friedrich Lewy) imposent :

$$\frac{\Delta t}{\Delta x}\leqslant\sqrt{\frac{\Delta t}{2a}} \tag{11.166}$$

d'où la condition sur Δt :

$$\Delta t\leqslant\frac{\Delta x^2}{2a} \tag{11.167}$$

et on retrouve une des conditions du paragraphe 3. Pour retrouver l'autre condition, dérivons l'équation initiale par rapport au temps, soit :

$$\begin{aligned}\frac{\partial}{\partial t}\left(\frac{\partial u}{\partial t}+\mathrm{V}\frac{\partial u}{\partial x}-a\frac{\partial^2 u}{\partial x^2}\right)&=\frac{\partial^2 u}{\partial t^2}+\mathrm{V}\frac{\partial^2 u}{\partial x\partial t}-a\frac{\partial^3 u}{\partial x^2\partial t}\\&=\frac{\partial^2 u}{\partial t^2}+\mathrm{V}\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial t}\right)-a\frac{\partial^2}{\partial x^2}\left(\frac{\partial u}{\partial t}\right)=0\end{aligned} \tag{11.168}$$

ce qui donne, en remplaçant $\frac{\partial u}{\partial t}$ par $-\mathrm{V}\frac{\partial u}{\partial x}+a\frac{\partial^2 u}{\partial x^2}$:

$$\frac{\partial^2 u}{\partial t^2}=\mathrm{V}^2\frac{\partial^2 u}{\partial x^2}-2a\mathrm{V}\frac{\partial^3 u}{\partial x^3}+a^2\frac{\partial^4 u}{\partial x^4} \tag{11.169}$$

ce qui permet d'expliciter la dérivée temporelle d'ordre 2 en fonction des dérivées spatiales. En reportant cette expression dans l'équation avec les termes de troncature, nous avons :

$$\frac{\partial u}{\partial t}+\mathrm{V}\frac{\partial u}{\partial x}-\left(a-\frac{\mathrm{V}^2\Delta t}{2}\right)\frac{\partial^2 u}{\partial x^2}=a\mathrm{V}\Delta t\frac{\partial^3 u}{\partial x^3}-\frac{a^2\Delta t}{2}\frac{\partial^4 u}{\partial x^4} \tag{11.170}$$

en négligeant les dérivées d'ordre 3 et 4, nous avons l'équation :

$$\frac{\partial u}{\partial t}+\mathrm{V}\frac{\partial u}{\partial x}-\left(a-\frac{\mathrm{V}^2\Delta t}{2}\right)\frac{\partial^2 u}{\partial x^2}=0 \tag{11.171}$$

qui est une équation d'advection diffusion dont le coefficient de diffusion a' est défini par $a'=a-\mathrm{V}^2\Delta t/2$. Ce coefficient ne peut être négatif, ce qui serait physiquement impossible ; cette considération impose donc la condition :

$$\Delta t\leqslant\frac{2a}{\mathrm{V}^2} \tag{11.172}$$

Classiquement ; le coefficient a' est appelé coefficient de diffusion numérique et fait donc intervenir la valeur du pas de temps Δt et la vitesse.

11.8.4 Équation d'advection

Afin d'illustrer ces phénomènes, on considère l'équation d'advection linéaire mono-dimensionnelle suivante :

$$\frac{\partial u}{\partial t}+\mathrm{V}\frac{\partial u}{\partial x}=0 \tag{11.173}$$

où la vitesse de convection V est connue et supposée positive.

Schéma FTCS

Si on applique le schéma FTCS vu au début de ce chapitre, nous aurons :

$$\frac{U_j^{n+1}-U_j^n}{\Delta t}+V\frac{U_{j+1}^n-U_{j-1}^n}{2\Delta x}=0 \tag{11.174}$$

soit encore :

$$U_j^{n+1}=U_j^n-\frac{C}{2}(U_{j+1}^n-U_{j-1}^n) \tag{11.175}$$

Il vient $a_j^+=-C/2$, $a_j^c=1$ et $a_j^-=C/2$. En calculant l'erreur de troncature, on peut montrer que ce schéma est consistant et que l'erreur de troncature est en $\mathcal{O}(\Delta t,\Delta x^2)$:

$$\frac{U_j^{n+1}-U_j^n}{\Delta t}+V\frac{U_{j+1}^n-U_{j-1}^n}{2\Delta x}=\frac{\partial u}{\partial t}+V\frac{\partial u}{\partial x}+\text{E.T.} \tag{11.176}$$

avec

$$\text{E.T.}=\frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2}+\frac{\Delta x^2}{2}\frac{\partial^3 u}{\partial x^3}+\mathcal{O}(\Delta t^2,\Delta x^3) \tag{11.177}$$

Le facteur d'amplification g est défini par $g=1-iC\sin\theta$ et il apparaît immédiatement que ce schéma est instable car $g>1$. Or pour l'équation d'advection diffusion, on avait vu que le schéma FTCS était stable sous conditions. On peut donc en déduire que ce sont les termes de diffusion qui permettent, sous conditions, d'assurer la stabilité. Pour obtenir un schéma stable, l'idée est donc d'introduire artificiellement dans le schéma numérique un peu de diffusion.

Méthode de Lax

Si on remplace, dans le schéma précédent ; U_j^n par $(U_{j+1}^n+U_{j-1}^n)/2$, on obtient le schéma de Lax qui s'écrit :

$$\frac{U_{j-1}^{n+1}-(U_{j+1}^n+U_{j-1}^n)/2}{\Delta t}+V\frac{U_{j+1}^n-U_{j-1}^n}{2\Delta x}=0 \tag{11.178}$$

soit encore :

$$U_j^{n+1}=\frac{1-C}{2}U_{j+1}^n+\frac{1+C}{2}U_{j-1}^n \tag{11.179}$$

avec $a_j^+=(1-C)/2$, $a_j^c=0$ et $a_j^-=(1+C)/2$. Le coefficient d'amplification est alors :

$$g=\cos\theta-iC\sin\theta \tag{11.180}$$

dont le module est $g^2=\cos^2\theta+C^2\sin^2\theta$ Pour que g soit inférieur à 1, il faut que C^2 soit aussi inférieur à 1, soit :

$$C=\frac{V\Delta t}{\Delta x^2}<1 \quad \text{si} \quad V>0 \tag{11.181}$$

L'erreur de troncature de ce schéma devient :

$$\frac{U_{j-1}^{n+1}-(U_{j+1}^n+U_{j-1}^n)/2}{\Delta t}+V\frac{U_{j+1}^n-U_{j-1}^n}{2\Delta x}=\frac{\partial u}{\partial t}+V\frac{\partial u}{\partial x}+\text{E.T.} \tag{11.182}$$

avec :

$$\text{E.T.} = \frac{\Delta t}{2} u_{tt} - \frac{\Delta x^2}{2\Delta t} u_{xx} + \frac{\Delta t^2}{3!} u_{ttt} + \frac{\Delta x^2}{3!} u_{xxx} \tag{11.183}$$

et on voit apparaître un terme en $\Delta x^2/2\Delta t$ et on doit donc assurer la convergence.

Si on utilise l'équation initiale :

$$\frac{\partial u}{\partial t} + \mathrm{V}\frac{\partial u}{\partial x} = 0 \tag{11.184}$$

en dérivant par rapport au temps, nous aurons :

$$\frac{\partial^2 u}{\partial t^2} + \mathrm{V}\frac{\partial^2 u}{\partial x \partial t} = 0 \tag{11.185}$$

soit :

$$\frac{\partial^2 u}{\partial t^2} = -\mathrm{V}\frac{\partial^2 u}{\partial x \partial t} = \mathrm{V}^2 \frac{\partial^2 u}{\partial x^2} \tag{11.186}$$

et en reportant dans l'erreur de troncature, nous avons :

$$\text{E.T.} = \frac{\partial u}{\partial t} + \mathrm{V}\frac{\partial u}{\partial x} - \frac{\Delta x^2}{2\Delta t}(1-\mathrm{C}^2) u_{xx} + \frac{\Delta t^2}{3!} u_{ttt} + \frac{\Delta x^2}{3!} u_{xxx} \tag{11.187}$$

et on peut interpréter cette équation sous la forme :

$$\frac{\partial u}{\partial t} + \mathrm{V}\frac{\partial u}{\partial x} - a' \frac{\partial^2 u}{\partial x^2} = 0 \tag{11.188}$$

qui est une équation d'advection diffusion discrétisée avec une erreur en $\mathcal{O}(\Delta t^2; \Delta x^2)$ avec une valeur de la diffusion a' définie par :

$$a' = \frac{\Delta x^2}{2\Delta t}(1-\mathrm{C}^2) \tag{11.189}$$

qui est positif si $\mathrm{C} \leqslant 1$.

Schémas décentrés

Au lieu d'utiliser une dérivée centrée pour les termes de convection ; on utilise une dérivée décentrée, en supposant que V est positif, soit :

$$\frac{\mathrm{U}_j^{n+1} - \mathrm{U}_j^n}{\Delta t} + \mathrm{V}\frac{\mathrm{U}_j^n - \mathrm{U}_{j-1}^n}{2\Delta x} = 0 \tag{11.190}$$

soit encore :

$$\mathrm{U}_j^{n+1} = (1-\mathrm{C})\mathrm{U}_j^n + \mathrm{C}\mathrm{U}_{j-1}^n \tag{11.191}$$

avec $a_j^+ = 0$, $a_j^c = 1 - \mathrm{C}$ et $a_j^- = \mathrm{C}$. Dans ces conditions :

$$g = (1 - \mathrm{C}(1-\cos\theta) - \mathrm{i}\mathrm{C}\sin\theta \quad \text{et} \quad g^2 = (1-\mathrm{C}(1-\cos\theta))^2 + \mathrm{C}^2\sin^2\theta \tag{11.192}$$

soit $g^2 < 1$ si $\mathrm{C} < 1$. En effet, si on applique les conditions de stabilité vues précédemment, alors :

$$(a_j^+ - a_j^-)^2 - (a_j^+ + a_j^-) \leqslant 0 \quad \text{et} \quad (a_j^+ + a_j^-)^2 - (a_j^+ + a_j^-) \leqslant 0 \tag{11.193}$$

il vient $C(C-1) \leqslant 0$, soit $C \leqslant 1$. Comme V est positif et que l'on prend les points j et $j-1$, on dit que la dérivée spatiale est prise « au vent » ou upwind. Si on avait choisi de prendre les points j et $j+1$, nous aurions eu :

$$\frac{U_j^{n+1}-U_j^n}{\Delta t}+V\frac{U_{j+1}^n-U_j^n}{\Delta x}=0 \tag{11.194}$$

et :

$$U_j^{n+1}=CU_{j+1}^n+(1+C)U_j^n \tag{11.195}$$

et par conséquent le module d'amplification :

$$g=1+C(1+\cos\theta)+iC\sin\theta \tag{11.196}$$

est toujours de module supérieur à 1.

Si V est négatif, on choisira la discrétisation suivante :

$$\frac{U_j^{n+1}-U_j^n}{\Delta t}+V\frac{U_{j+1}^n-U_j^n}{\Delta x}=0 \quad \text{et} \quad U_j^{n+1}=CU_{j+1}^n+(1+C)U_j^n \tag{11.197}$$

où C est négatif. On montre facilement que dans ce cas, le schéma est stable si $|C|<1$.

Calculons l'erreur de troncature dans le cas où V est positif avec un schéma upwind :

$$\frac{U_j^{n+1}-U_j^n}{\Delta t}+V\frac{U_j^n-U_{j-1}^n}{\Delta x}=\frac{\partial u}{\partial t}+V\frac{\partial u}{\partial x}+\text{E.T.} \tag{11.198}$$

avec :

$$\text{E.T.}=\frac{\Delta t}{2}u_{tt}-\frac{V\Delta x}{2}u_{xx}+\frac{\Delta t^2}{3!}u_{ttt}+\frac{V\Delta x^2}{3!}u_{xxx} \tag{11.199}$$

et on montre que le schéma est consistant et d'ordre $(\Delta t,\Delta x)$. Comme dans le paragraphe précédent ; on peut remplacer u_{tt} par V^2u_{xx} et écrire le résultat précédent sous la forme :

$$\begin{aligned}\text{E.T}&=\left(\frac{V^2\Delta t}{2}-\frac{V\Delta x}{2}\right)u_{xx}+\frac{\Delta t^2}{3!}u_{ttt}+V\frac{\Delta x^2}{3!}u_{xxx}\\&=\frac{V\Delta x}{2}(C-1)u_{xx}+\frac{\Delta t^2}{3!}u_{ttt}+V\frac{\Delta x^2}{3!}u_{xxx}\end{aligned} \tag{11.200}$$

et le coefficient de diffusion numérique est $a'=(1-C)V\Delta x/2$ qui est positif si $C<1$.

Schéma de Lax-Wendroff

Dans ce schéma, on se fixe la valeur de la diffusion numérique de façon à ce que le coefficient de la dérivée seconde en espace soit identiquement nul dans l'erreur de troncature, un fois la dérivée seconde en temps remplacée par son expression en fonction de la dérivée spatiale. Ce schéma s'écrit sous la forme suivante :

$$\frac{U_j^{n+1}-U_j^n}{\Delta t}+V\frac{U_{j+1}^n-U_{j-1}^n}{2\Delta x}-\frac{V^2\Delta t}{2}\frac{U_{j+1}^n-2U_j^n+U_{j-1}^n}{\Delta x^2}=0 \tag{11.201}$$

soit sous la forme adimensionnelle :

$$U_j^{n+1}=\frac{C(1+C)}{2}U_{j-1}^n+(1-C^2)U_j^n-\frac{C(1-C)}{2}U_{j+1}^n \tag{11.202}$$

avec $a_j^+ = -\mathrm{C}(1-\mathrm{C})/2$, $a_j^c = 1-\mathrm{C}^2$ et $a_j^- = \mathrm{C}(1+\mathrm{C})/2$. Le coefficient d'amplification s'écrit :

$$g = 1-\mathrm{C}^2(1-\cos\theta)-\mathrm{iC}\sin\theta \tag{11.203}$$

ce qui conduit à la condition $\mathrm{C} \leqslant 1$.

On vérifie que l'erreur de troncature est de la forme :

$$\text{E.T.} = \frac{\Delta t}{2}\frac{\partial^2 u}{\partial t^2} - \frac{\mathrm{V}^2\Delta t}{2}\frac{\partial^2 u}{\partial x^2} + \frac{\Delta t^2}{6}\frac{\partial^3 u}{\partial t^3} + \frac{\mathrm{V}\Delta x^2}{6}\frac{\partial^3 u}{\partial x^3} + \mathcal{O}(\Delta t^3, \Delta x^3) \tag{11.204}$$

en remplaçant $\frac{\partial^2 u}{\partial t^2}$ par $\mathrm{V}^2\frac{\partial^2 u}{\partial x^2}$ et $\frac{\partial^3 u}{\partial t^3}$ par $-\mathrm{V}^3\frac{\partial^3 u}{\partial x^3}$, il vient :

$$\text{E.T.} = \left(\frac{\mathrm{V}\Delta x^2}{6} - \frac{\mathrm{V}^3\Delta t^2}{6}\right)\frac{\partial^3 u}{\partial x^3} + \mathcal{O}(\Delta t^3, \Delta x^3) = \frac{\mathrm{V}\Delta x^2}{6}(1-\mathrm{C}^2)\frac{\partial^3 u}{\partial x^3} + \mathcal{O}(\Delta t^3, \Delta x^3) \tag{11.205}$$

11.9 Coordonnées cylindriques ou sphériques

On considère maintenant une équation aux dérivées partielles du second ordre de type parabolique où les deux variables sont le temps t et une dimension d'espace r (coordonnées cylindriques ou sphériques). Soit $u(r,t)$ cette fonction et l'opérateur différentiel $\mathscr{L}$ qui est défini par :

$$\mathscr{L}(u) = \frac{\partial u}{\partial t} + \mathrm{V}\frac{\partial u}{\partial r} - a\left(\frac{\partial^2 u}{\partial r^2} + mr\frac{\partial u}{\partial r}\right) \tag{11.206}$$

avec :

- $m = 1$ en coordonnées cylindriques ;
- $m = 2$ en coordonnées sphériques.

où on suppose que V et a sont des constantes ou des fonctions du temps t et de la variable d'espace r. Les équations (11.207) et (11.208) correspondent en fait à des problèmes possédant une symétrie. Le problème majeur est que ces équations dégénèrent en $r = 0$:

$$\mathscr{L}(u) = \frac{\partial u}{\partial t} + \mathrm{V}\frac{\partial u}{\partial r} - a(1+m)\frac{\partial^2 u}{\partial r^2} \tag{11.207}$$

Soit à résoudre le problème suivant :

$$\begin{array}{l} \mathscr{L}(u) = 0 \text{ pour } t > 0 \text{ et } r \in [r_a, r_b] \\ u(0,r) \text{ donné : condition initiale} \\ r = r_a \text{ et } r = r_b \text{ en fonction du temps : conditions limites} \end{array} \tag{11.208}$$

RÉFÉRENCES

[Bettess 92] Bettess P., *Infinite elements*, Penshaw Press, 1992.

[Bonnet 95] Bonnet M., *Équations intégrales et éléments de frontière. Applications en mécanique des solides et des fluides*, CNRS Éditions/Eyrolles, 1995.

[Brebbia 78] Brebbia C.A., *Recent advances in boundary elemnt methods*, Pentech Press, 1978.

[Euvrard 90] Euvrard D., *Résolution numérique des équations aux dérivées partielles. Différences finies, éléments finis, méthode des singularités*, Masson Éditions, 1990.

[Wielgosz 82] Wielgosz C., Informations exactes données par des méthodes d'éléments finis en mécanique, vol. 1(2), 1982, pp. 323–329. 25

[Wielgosz 99] Wielgosz C., *Cours et exercices de résistance des matériaux. Élasticité, plasticité, éléments finis*, Éditions Ellipses - Marketing, Paris, 320 pages, 1999. 5, 10, 25